AGRICULTURAL RESEARCH INSTITUTE

PUSA

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

KARL PEARSON

ASSISTED BY

EGON SHARPE PEARSON

VOLUME XXIV

1932

ISSUED BY THE BIOMETRIC LABORATORY
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

CONTENTS OF VOLUME XXIV

Memoirs

		PAGE
I.	Hereditary Entropion and Hereditary Changes in the Skin of the Eyelids. By C. H. USHER. With fourteen figures in the text, one folding pedigree and four plates	1
II.	The Geometric Properties of Microscopic Configurations. I. General Aspects of Projectometry. By WILLIAM R. THOMPSON . .	21
III.	The Geometric Properties of Microscopic Configurations. II. Incidence and Volume of Islands of Langerhans in the Pancreas of a Monkey. By WILLIAM R. THOMPSON, RAYMOND HUSSEY and others. With one figure in the text	27
IV.	A Bessel Function Distribution. By A. T MCKAY. With one diagram in the text	39
V.	On the Normality or Want of Normality in the Frequency Distributions of Cranial Measurements. By E. M. ELDERTON and T. L. WOO. With three diagrams in the text	45
VI.	The Sampling Distribution of the Third Moment Coefficient—an Experiment. By JOSEPH PEPPER. With three figures in the text	55
VII.	Further Contributions to the Sampling Problem. By N. ST GEORGESCU	65
VIII.	A Preliminary Classification of Asiatic Races based on Cranial Measurements. By T. L. WOO and G. M. MORANT. With one map, two diagrams in text and two folding sheets of tables .	108
IX.	A Study of the Crania in the Vaulted Ambulatory of St Leonard's Church, Hythe. By BRENDA N. STOESSIGER and G. M. MORANT. With one map, one plan of Hythe, and six contours in text. Six tissue contours in pocket, two folding sheets of measurements and seven plates	135
X.	On the Mean Character and Variance of a Ranked Individual, and on the Mean and Variance of the Intervals between Ranked Individuals. By KARL PEARSON, with the assistance of MARGARET V. PEARSON. Part II. The Case of Certain Skew Curves. With three diagrams and eleven figures in the text	203

		PAGE
XI.	Further Applications in Statistics of the $T_m(x)$ Bessel Function. By KARL PEARSON, S. A. STOUFFER, and F. N. DAVID. With three figures in the text	293
XII.	Experimental Discussion of the (χ^2, P) Test for Goodness of Fit. By KARL PEARSON	351
XIII.	On the Distribution of the Correlation Coefficient in Small Samples. By PAUL R. RIDER. With six figures in the text	382
XIV.	The Percentage Limits for the Distribution of Range in Samples from a Normal Population. By EGON S. PEARSON. With one figure in the text	404
XV.	The Converse of Spearman's Two-factor Theorem. By BURTON H. CAMP	418
XVI.	The Distribution of the Index in a Normal Bivariate Population. By E C. FIELLER. With two figures in the text	428
XVII.	A Note on the Distribution of the Correlation Ratio. By JOHN WISHART	441
XVIII.	On the Probability that Two Independent Distributions of Frequency are really Samples from the Same Parent Population. By KARL PEARSON	457
XIX.	Certain Generalizations in the Analysis of Variance. By S. S. WILKS	471

Miscellanea

(i)	A Comparison of the Accuracy of Two Types of Quadrature Formulae. By E. S. MARTIN	280
(ii)	A Simple Non-Normal Correlation Surface. By H. L. RIETZ	288
(iii)	Note on Professor Rietz's Problem. EDITORIAL	290
(iv)	Note on a Memoir by A. E. R. Church, *Bi*[illegible] Vol. XVIII. By A. E. R. CHURCH	292
(v)	On a Method of Proceeding from partial Cell Frequencies to Ordinates and to total Cell Frequencies in the case of a bivariate Frequency Surface. By JACQUES CHAPELIN	495
(vi)	On the Betas of Quadrilateral Distributions. By OWEN L. DAVIES. With three figures in the text	498

List of Plates, etc.

Biometrika Portrait Series, No. IX. A. A. MARKOFF, 1856—1922 *Frontispiece*

(α) *C. H. Usher:* Hereditary Entropion and Hereditary Changes in the Skin of the Eyelids.

Folding Pedigree . . . . *to face page* 10

Plates I—IV . . . . „ „ 20

(β) *T. L. Woo and G. M. Morant:* A Preliminary Classification of Asiatic Races based on Cranial Measurements.

Folding Table of Coefficients of Racial Likeness between Oriental Races . . . . „ „ 122

Folding Table of Coefficients of Racial Likeness between Northern Mongolian and Oriental Races . . . . . „ „ 126

(γ) *B. N. Stoessiger and G. M. Morant:* A Study of the Crania in the Vaulted Ambulatory of Saint Leonard's Church, Hythe.

Plate I. Engravings of the Interior of the Ambulatory Passage of Saint Leonard's Church, Hythe . „ „ 202

Plate II. Typical Hythe Skulls. *Norma lateralis* . . „ „ „

Plate III. „ „ „ *Norma facialis* . . „ „ „

Plate IV. „ „ „ *Norma verticalis* . . „ „ „

Plate V. Abnormal Hythe Skulls . . . . „ „ „

Plate VI. Basal Anomalies of the Hythe Skulls . . . . „ „ „

Plate VII. Anomalous Hythe Skulls . . . . „ „ „

Appendix of Individual Measurements in three tables on two folding sheets . . . . „ „ „

Six Cranial Contours on tissue in pocket at end of volume.

A. A. Markoff, 1856–1922.

"*Biometrika*" *Portrait Series*, No. IX. Issued as frontispiece to Vol. XXIV.

VOLUME XXIV MAY, 1932 PARTS I AND II

BIOMETRIKA

HEREDITARY ENTROPION AND HEREDITARY CHANGES IN THE SKIN OF THE EYELIDS.

BY C. H. USHER, M.B., B.CH.

THE entropion in this pedigree is of the spasmodic variety and appears at a much earlier age than the usual spasmodic entropion of elderly people. A few of the cases are accompanied by obvious changes in the skin of the lids, but these changes in the skin occur more often without entropion.

Before describing these cases a short note will be given, more particularly on the clinical aspect of cases in the literature that appear to have a bearing on the subject of these changes in the skin of the lids, and some notes on congenital entropion.

Changes in the Skin of the Lids. Graf, in 1836, under the title "Local hereditary relaxation of the skin," described a case in which both lower lids were like sacks, with skin thinned, in longitudinal and transverse folds; the left side of the neck was also affected. Sichel, in 1844, describes 'ptosis atonique': "The skin of the upper lids is flask shaped, wrinkled, sometimes even folded transversely, it may sometimes hang in front of the tarsus in form of a transverse fold which descends to its free edge, loss of elasticity is noted, the head is carried erect to enable the patient to see." Mackenzie (1854) says, "It is occasionally the case, that when the fold of integument is very considerable it presses by its weight, the edge of the lid, along with the cilia, inwards, so as to produce a degree of entropium." Arlt (1874) stated that ptosis adiposa is seen in young people in whom the fold of the skin on the upper eyelids is so enlarged especially at the outer angle that it lies over the lashes; the skin is red, but not inflamed. If a fold of this thin and flexible skin is incised, a soft yellow elastic fat is found immediately beneath, which is indistinguishable from orbital fat tissue, and is so pressed forwards that a considerable portion of it requires to be removed with scissors in order to close the skin wound satisfactorily. Fuchs (1896) described a group of cases as blepharochalasis. The disease affects exclusively the upper lids. The clinical signs are these: Both upper lids are affected, skin is very thin and has lost elasticity, in consequence fine folds lie in all directions, as is seen in extensive senile atrophy and relaxation of the skin. The skin resembles crinkled cigarette paper. Increase in surface of skin, peculiar reddening of lids, numerous small distended veins in the skin like those seen in old people with red cheeks. Skin changes are most marked between brow and upper edge of tarsus. Relaxation of the subcutaneous cellular tissue. As a result the skin hangs down in the form of an unpleasant relaxed and reddened sack over the lid margin and may suggest ptosis. Occurs in male and female, in youth and middle age. The patients call the condition

a swelling. In most cases it could not be determined whether the change in the lids had been preceded by oedematous swelling or not. Fuchs mentions that all these cases of drooping of the skin over the lid margin had been given the common name epiblepharon (v. Ammon). Since this paper was published a number of cases have been described bearing on the subject of blepharochalasis. Fehr (1898), in his case of "Lidhaut-Erschlaffung sog. Blepharochalasis," makes no mention of wrinkling or of veins in the skin of the lids. Schmidt-Rimpler (1899), in his paper on "Fett-Hernien der oberen Augenliden," describes a protrusion of fat through the musculature of the lid found at operation. A similar case was reported by Meguro in 1930. Rohmer (1900), in an article "De l'angio-mégalie symétrique des paupières supérieures," describes four cases confined to the upper lids of women. Terson (1904) describes three cases under the heading "Dermatolysie palpébrale." In the first case the upper lids were swollen and hyperaemic. At times the thin skin hung as a sack on the lashes. In the second case the upper lids were swollen and red. In the third case the skin of the left upper lid showed since birth expansion and an unsightly droop. Scrini's case (1906) had the upper lids swollen from oedema during migrainous attacks. The orbital lachrymal gland was felt through the outer part of the orbito-palpebral furrow. The free edge of the lid in two-thirds of its extent was covered by a cutaneous fold formed bv flabby skin, thin as honeycomb, and of pale colour. With cold it became red, some venules seen especially in the left one. Lafon and Villemont (1906) describe a case on the upper lid showing a continuous fold parallel to the ciliary margin; at the junction of inner and middle third of the lid this fold descended obliquely downwards and outwards, forming an apron which hid the ciliary margin and the upper part of the outer canthus; the pupil was partly hidden. A hard, painless, mobile body was felt in the fold, an ectopic orbital lachrymal gland. No wrinkling of the skin is mentioned. No oedema or congestion of the lids had been noticed until a recent attack in the left. Watering occurred readily from wind and cold. The title of the paper is "Blépharochalasis Héréditaire avec Dacry-adénoptose." Bach (1906) believes that the clinical picture of blepharochalasis may be produced by a collection of fat in the upper lids. Wagenmann (1907) reported a case of ptosis adiposa in a youth of 16. Weinstein (1909) records a typical case of ptosis adiposa that is, he says, a case of ptosis atonica complicated by the occurrence of fat under the skin. Ptosis atonica and ptosis adiposa appear as two different forms of the same disease. He would prefer the designation ptosis atrophica for ptosis atonica. Loeser (1908) found luxation of the lachrymal gland in a case of blepharochalasis. Stieren (1914) reported two cases of blepharochalasis. In the first subcutaneous masses due to fat were present and could be reduced into the orbit, when the appearances were those of wrinkled skin as described by Fuchs. In the second case there were several attacks of swelling of lids, the skin was wrinkled and thin, no masses were felt in the eyelids, but a rounded elongated mass was felt under the upper margin of orbit. In Jenison's case of blepharochalasis (1915) swelling was not intermittent. Skin of upper lids was thin, shiny, pink and there was definite atrophy. Dilated veins very evident, lachrymal glands not palpable, no excess of fat in upper lids. Wassermann test; given Hg. bichloride and potas. iodide. Two days later pain

in lids followed by swelling. On 10th day swelling had disappeared but skin hung lax and flaccid. Heckel's (1921) case designated blepharochalasis with ptosis was unilateral. Verhoeff and Friedenwald (1922), in a paper on blepharochalasis, describe the case of a female with upper lids swollen, skin finely wrinkled, loose and redundant like that of old age when fat tissue has disappeared, but in contrast to this lids were puffy. No nodules present. Friedenwald (1923) describes the case of a male with both upper lids baggy, twice as much skin surface as is normal, skin soft and pliable, in no way suggesting oedema, with many fine wrinkles and deficiency of subcutaneous connective tissue. The excess skin formed a fold which extended a little below ciliary margin and covered upper half of pupillary areas. Cilia of outer one-third of upper lids were turned in and constantly rubbed against the cornea. Weidler's first case (1913) had bagginess and drooping of skin and subcutaneous tissue over edge of lid margin. No wrinkling or folding, skin smooth. In second case the skin, pinkish red, hung down in a baggy pouch-like mass partially covering the eyes, superficial veins prominent, no wrinkling mentioned, lachrymal gland could be rolled under the finger. Benedict (1926) describes three cases of blepharochalasis.

Attention has been drawn by a few writers in Germany and Holland to the association between the occurrence of double lip* and changes in the skin of the lids. Ascher (1920) describes several cases with blepharochalasis, goitre and double lip. Wirths (1920), in a paper entitled "Beiderseitige Lidgeschwulst, kombiniert mit Geschwulstbildung der Oberlippe," gives the case of a male, age 23, with a history of changes in both lids and upper lip since birth. Weve (1921) described the case of a soldier, age 27, with typical blepharochalasis and double lip. Eigel's case (1925) was that of a man who according to his mother developed double lip and blepharochalasis at the age of 10.

The association of double lip with blepharochalasis has no immediate concern with the subject of this paper, for no double lip has been found in any member of the pedigree. But as these cases are not well known, notes of two examples, shown to me recently by medical friends, may not be out of place here.

Case 1. (Plate I, Fig. 1.) D. Mc.G., male, age 74, seen 10th May, 1931. On the upper lids are large folds of skin. These cover the whole of the lid margin so that no lashes of the upper lid are seen on the left side and only the ends of a few on the right side. The edge of the fold covers the upper part of the pupil, so that when talking he holds his head well back to enable him to see better. The skin of the

* The term "double lip" is applied to a condition occasionally met with in young men, in which there is an hypertrophy of the labial glands in the mucous membrane of the upper lip. It is of slow growth, and forms an elongated swelling on each side of the frenum, covering the teeth and projecting the lip. It is shotty to the feel, and the only complaint is of disfigurement. The treatment consists in excising the redundant fold of mucous membrane, including the enlarged mucous glands. *Manual of Surgery*, by Alexis Thomson and Alexander Miles, Vol. 2, 5th edition, p. 165, 1915.

Neustätter distinguished two kinds of double lip in adults. See Ascher's first paper. In a treatise on surgery by George Ryerson Fowler, Vol. 1, p. 477, 1906, occur the following words: "Congenital hyperplasia of the labial substance is sometimes observed. The thickening may be due to an excessive thickening of lymph-vessels (lymph-angioma) or the hyperplastic condition may refer more to the mucous membrane, becoming visible as a 'double lip' during the act of laughing."

fold is very thin and has numerous fine wrinkles placed in all directions. There is marked loss of elasticity of the skin, which shows numerous blue veins. Lower lids have wrinkled skin, but otherwise are normal. There is a horizontal, smooth, soft ridge behind his upper lip. This has the appearance and consistence of mucous membrane. It obscures his upper gums with his few remaining teeth. His thyroid gland is not enlarged. He dates the lid condition from the age of seven or eight, and at the age of 19 the unusual appearance was commented on by the late Sir William McGregor.

Case 2. (Plate II, Fig. 2.) Mr G. MacD., age 64, when looking horizontally forwards, nearly the whole of the margin of the left upper lid is covered by a fold of skin and more than half of the right upper lid margin is similarly covered. There is moderate wrinkling of the skin fold, no veins are visible in the skin. He noted the folds for the first time two years ago. Marked double lip which is soft and red like mucous membrane. It extends laterally nearly to angles of mouth where it blends with the proper upper lip. The frenum joins the mucous membrane of the double lip. He has had the condition since boyhood. His natural front teeth are present. He knows of no one of his own family with the same condition. To illustrate marked wrinkling of the skin of the lids and loss of elasticity Fig. 3 is shown. The two vertical ridges on the left upper lid have remained after pinching up the skin. The photograph is that of a female, age 57. The condition began to develop when she was 21. She also states that her mother's eyelids were similarly affected.

From the literature it is obvious that skin folds of the lids have received a number of names by different writers, and it is not always possible to decide whether the conditions referred to by some under one name are the same as those described by others under other names. Scrini regards his case as the blepharochalasis of Fuchs, the cutaneous ptosis of Panas, the paupière en besace of Frenkel, the angiomégalie palpébrale of Rohmer or yet the dermatolysie palpébrale, but with subluxation of the orbital lachrymal gland. Elschnig, according to Ascher, reserves the term blepharochalasis for those cases which besides skin atrophy and laxness show, at least with pressure on the eyeball, orbital contents (fat, lachrymal gland). He excludes, therefore, atrophy of the skin of the lid in the aged, lid skin atrophy with laxness and drooping, the "paupière en besace" according to Ginestous and Frenkel, ptosis atrophia according to Weinstein, and many cases of the old ptosis atonica. On the other hand he includes fat hernia of the upper lid (Schmidt-Rimpler), lipomatosis of upper lid and ptosis adiposa, which Wagenmann includes as hypertrophy of the fat tissue. Fuchs expressly states that he does not wish blepharochalasis to be included in the name ptosis atonica.

As the skin changes in the present cases take several forms it was deemed unsatisfactory to attempt to include them under any less comprehensive heading than "changes in the skin of the lids." The term epiblepharon or ptosis atonica would be suitable for some of the forms and possibly blepharochalasis in a few cases, though none of these cases agree exactly with Fuch's description.

Microscopical examination of the upper lid tissue in cases with skin folds has been made by a number of observers. Atrophic skin changes were found by some

and infiltration by others. Elastic fibres have been found diminished by some investigators and normal by others. In blepharochalasis Miglietta found vascular and lymphatic changes in the skin of the lids. As none of the cases in this pedigree were examined microscopically, it is not intended to pursue the subject further. Reports have been published by Schmidt-Rimpler, Rohmer, Rosenstein, Lodato, Scrini, Stieren, Heckel, Wagenmann, Verhoeff and Friedenwald, Friedenwald, Eigel, Weidler, Miglietta and others.

The age at onset of blepharochalasis has been given from infancy to 20 years. It is apparently an uncommon condition. Ascher states that only one case was seen among 30,000 out-patients. On the other hand Lafon and Villemont consider that it is not so rare as Scrini's article might make one think, and say that all ophthalmologists have observed it, but that in general the patients are not much concerned about an infirmity so trifling and they come to consult for other reasons.

The *etiology* of blepharochalasis has been ascribed to angioneurotic oedema of the upper lids, to fat from the orbit, and to a collection of fat in the upper lids. Rohmer believes that excess of orbital fat may have an influence at least on the degree of prominence of the skin of the lids, that dilated lymphatics are secondary, that there is an anomaly in development of the vascular system of the lids, and that the origin is undoubtedly vaso-motor, therefore his designation "angio-mégalie des paupières supérieures." Fuchs believed the cause was of neuropathic nature in some cases. Terson believes that the trophic disturbance in dermatolysie palpébrale probably depends on a neuro-trophic lesion dependent on the sympathetic. Scrini is inclined to agree with Terson that dermatolysie palpébrale depends on general causes and mentions digestive disorders, migraine accompanied by oedema of the upper lids, of which the integument by reason of congenital defect of tonicity becomes thinned following repeated distension and loses all its elasticity. Disordered menstruation is considered to be a causal factor by some. It has been suggested, says Benedict, that blepharochalasis may have some relation to the development of the thyroid. The onset at 10—17 years places it definitely within the period of functional endocrine development. Accardi also believes that as in the essential cutaneous atrophies of the dermatologists, it is necessary in blepharochalasis to ascribe a rôle to endocrine disturbance without being able to specify which gland is at fault. Miglietta thinks that blepharochalasis is brought about by hormonal influence associated, or not, with dysfunction of the neurovegetative system on the one hand and familial predisposition on the other*. Lafon and Villemont suggest that atrophy and looseness of the suspensory fibres of the skin of the lid may be hereditary. Ptosis adiposa, according to Weidler, presents some of the signs of blepharochalasis. There is relaxation of the skin, but no true atrophy follows the condition. The bagging of the upper lid is more marked at the inner side in ptosis adiposa and is thought to be due to relaxation of bands of fascia connecting the skin with the tendons of the

* The author recalls the case of a female with atrophic striae on the thighs and under the breasts though she had never had children and had experienced no noteworthy thinning of the body pointing to destruction of elastic fibres of the same nature and determined by the same cause which acted on the eyelids.

levator and with the margin of the orbit. Mention has already been made of the conditions Elschnig includes and excludes under the name blepharochalasis.

Heredity appears in only a small number of the published cases of blepharochalasis. In Graf's case (Fig. 1), which Franceschetti apparently accepts as one of blepharochalasis, only the lower lids were affected. The family was Russian. The four cases in three generations were similarly affected. II, 1 married and had no issue. II, 2 was not affected and died in his 71st year. In the four cases the symptoms arose in the fourth decade. The cases showed also relaxation of the skin on left side of neck.

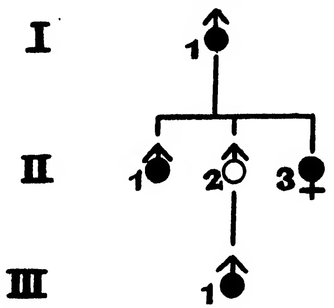


Fig. 1.

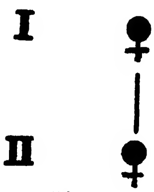


Fig. 2.

Schmidt-Rimpler's case of fat hernia of the upper lid in a female had a mother affected with the same condition (Fig. 2).

In Lafon and Villemont's case (Fig. 3) blepharochalasis occurred in four generations. The patient III, 1 was a man aged 43. His mother and maternal grandfather had a similar palpebral malformation. Besides he had five children of whom the second, a son aged 18, had the characteristic fold. The four other children were not affected. On the left side III, 1 had an ectopic lachrymal gland*.

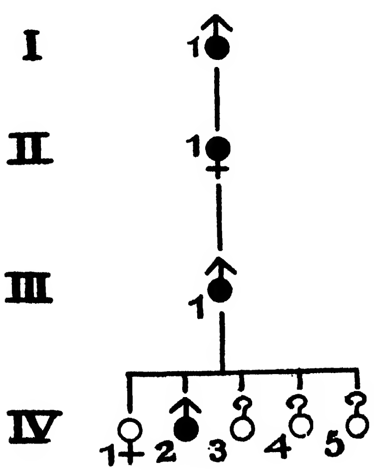


Fig. 3.

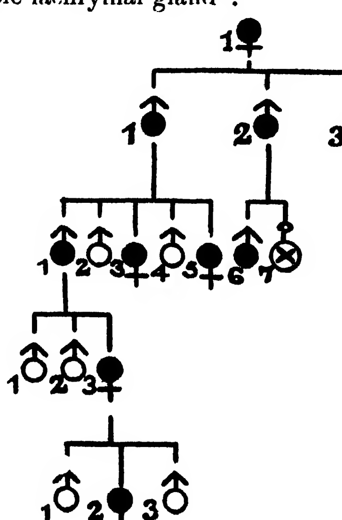


Fig. 4.

* See note on p. 5.

Franceschetti after referring to the rarity of hereditary blepharochalasis says that the skin fold of the upper lid that resembles blepharochalasis (epiblepharon) occurs as a familial form more frequently. In his Fig. 37* (Fig. 4 above) he shows a family tree. There is no description of the cases.

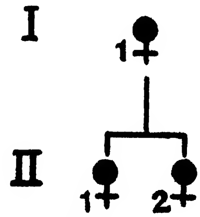


Fig. 5.

In Müller's first pedigree (Fig. 5) II, 1, age 1½, has a fold of skin on both lower eyelids and entropion of the right lower lid. I, 1 has a slight degree of epiblepharon of both lower lids. II, 2, age six months, has marked epiblepharon of both lower lids. The father of I, 1 according to I, 1 had peculiar lower lids, but it is not certain that he had epiblepharon.

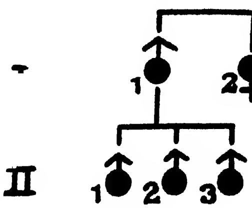


Fig. 6.

In Müller's second pedigree (Fig. 6) a brother and sisters and three sons of the brother had epiblepharon of both lower lids. II, 3, age seven, according to the father, had at times the lashes of the lower lids touching the eyeball. On present examination it could only be made out that the lashes of the lower lids were pressed obliquely upwards by the skin fold. In both adults the appearance of the lids corresponded with Elschnig's Fig. 1†.

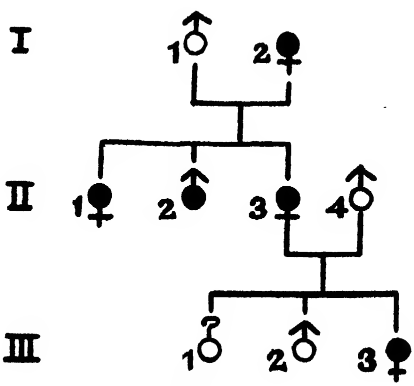


Fig. 7.

Miglietta's first Pedigree (Fig. 7). II, 3, a woman age 51, had blepharochalasis. I, 2, her mother, had similar folds in upper eyelids as patient. II, 1, a sister, and II, 2, a brother, had bilateral blepharochalasis from an early age, especially the brother whose affection had increased with age. The patient as a girl had enlarged cervical glands. She had chilblains on hands and feet. Menstruation irregular. Married and had three children. One died in infancy. III, 3, a daughter, has signs of blepharochalasis on left side.

The alteration of the eyelids dates from a young age. Patient has never noticed any swelling or reddening of the skin of the eyelids. The folds of skin increased more and more. The skin fold is pale, thin and inelastic, no net-work of veins is seen.

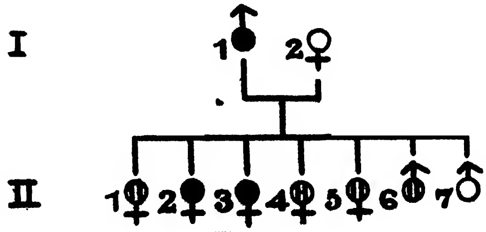


Fig. 8.

Miglietta's second Pedigree (Fig. 8). II, 2, a widow, age 62, has blepharochalasis. Her father I, 1, a small farmer, with blepharochalasis, died aged 68 of lung trouble. Patient has two brothers and four sisters. Almost all of these have an evident degree of blepharochalasis. Patient is the second born. She married

* "Abb. 37. Vererbung von Blepharochalasis bzw. Epiblepharon des Oberlides (Eigene Beobachtung)."

† *Med. Klinik* 1922, Jahrg. 18, nr. 16, S. 498.

and had an abortion. In the chart the only one in the sibship that is not shaded at all is II, 7.

Congenital Entropion. For readers unfamiliar with the subject it may be stated that entropion is a rolling inwards of the lid constituting trichiasis so that the lashes rub against the cornea. There is a cicatricial variety due to contraction of the conjunctiva as a result of burns, accidents, trachoma and other causes and a spasmodic or muscular variety caused by contraction of the orbicularis muscle. Himly states that late in foetal life the lid margins are markedly turned inwards, and that if this condition persists beyond the normal a slight trace of congenital entropion may result, but true entropion does not occur congenitally. In 1841 v. Ammon saw entropion of the lids of the left eye and of the upper lid of the right eye in a girl of three, also ectropion of the right lower lid. Primary congenital entropion is a rare occurrence and, according to Aubineau, de Wecker in the course of his long career had never met the deformity. Berry regards this variety of entropion as probably due to abnormal development of the orbicularis in the vicinity of the lid margin. Secondary entropion is associated with anomalies of the eyeball as microphthalmos and some conditions of the lids as epicanthus. There appear to be two forms of primary congenital entropion, (1) a spasmodic variety, and (2) a variety in which the tarsus is absent. Guibert (1892), in a case of double congenital entropion which the mother had noticed eight days after the birth of the child, found on operating absence of the tarsal cartilage. Harlan (1895) examined two cases of congenital entropion of both upper lids with deficiency of tarsal cartilages. Leblond (1907) mentions two cases of bilateral entropion in sisters. In the younger sister the inturning of the lids was noticed by the parents at birth. There was no epicanthus, skin was not hypertrophied. At operation the tarsus was made out to be well developed and there was excessive development of the orbicularis in the neighbourhood of the ciliary region. Polstoocchow (1914) saw two sisters with congenital entropion, which the mother had noticed immediately after birth. The mother also affirmed that her last child, aged some months, showed the same defect of development. Polstoocchow ascribed the pathogenesis to development of the ciliary bundles of the orbicularis muscle. In Sziklai's case (1917), a boy of four, the congenital entropion was due to an exaggerated development of the palpebral portion of the orbicularis muscle. In Hessberg's two cases (1922) the tarsus was normal. Yan Chow (1925) describes three cases of entropion of the new born operated on with success. The tarsus was in all cases normal. In his first case the upper lid was turned in. Schorr (1926) operated for spasmodic entropion of upper lid in a suckling. She gives the literature of spasmodic entropion of the upper lid. Aubineau (1928) records two cases of congenital entropion with spasmodic appearance and no deficiency of tarsus. He also records three cases of congenital entropion (in mother, son and daughter) from deficiency or absence of tarsus. Denig (1899) found in a child the inner part of the upper lid inverted. In two cases aged seven years and ten months respectively Dimmer (1885) attributes the entropion to excess of skin of the lower lid. Gomez Marquez (1930) reports congenital entropion of the lower lids in two cases. In one it was bilateral, in the other unilateral.

Hereditary congenital entropion is evidently very rare. The cases described by Leblond, Hessberg, Polstoocchow, and Aubineau are the only instances known to me, unless there is included Sydney Stephenson's cases (1894) of two brothers (Fig. 9), aged seven and six respectively, with congenital trichiasis, in which the intermarginal space was almost normally directed. Inheritance of congenital entropion has been

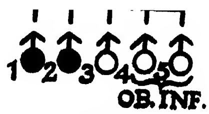


Fig. 9.

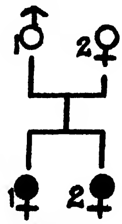


Fig. 10.

recorded by Leblond (1907) (Fig. 10). Of two sisters affected, the younger, age eight, had bilateral entropion of her lower lids. The inturning of the lids was noticed by the parents at birth. The elder sister, age 18, was operated on for an identical malformation at the age of four. Father and mother normal.

Polstoocchow (1914) saw (Fig. 11) two sisters with congenital entropion; in one of them it was bilateral, in the other unilateral. According to the mother, the condition was present from birth. She said that her youngest child, age some months, showed the same defect.

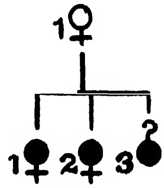


Fig. 11.

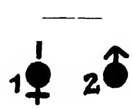


Fig. 12.

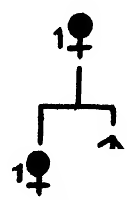


Fig. 13.

Hessberg (1922) recorded congenital familial entropion of both lower lids in a sister, age three, and brother, age two (Fig. 12).

Aubineau (1928) saw a mother, daughter and son with congenital entropion (Fig. 13). In the mother it was localised in the left lower lid. Both lower lids were affected in daughter and son. The mother was operated on. There was deficiency or absence of the tarsus.

Attention was drawn to the pedigree (see folding sheet) that is the subject of this paper by the occurrence of spasmodic entropion in two men at an unusually early age. Each gave a history of having relatives affected with the same condition. These two men proved to be first cousins. Investigation showed the correctness of their story and that in some cases the entropion was associated with unusual conditions of the skin of the lids, which also appeared in some cases apart from entropion. They were both in-patients, in 1896, at the Aberdeen Royal Infirmary.

David R. (III, 53, Plate I, Fig. 4), age 19 (Nov. 28, 1896), fisherman. **Diagnosis:** Congenital entropion of right lower lid. Right lashes turned in along whole length of lower lid and rubbing against the eyeball, cornea clear, refraction myopic, pupils, equal, contract to light, considerable photophobia; left lower lid tends to turn in when he is lying down, refraction of left eye hypermetropic. His paternal grandfather (I, 3) had both lower lids turned in. This is said to have been since birth. A brother (III, 51) had both lower lids turned in since birth. They were remedied by operation. A male (III, 39), first cousin, son of an uncle, had both lower lids turned in from birth and another first cousin (III, 41) had one lid turned in from birth. Patient has had excellent health. He comes because of defective vision, "sight goes away when at sea." He can usually see to work quite well, but in bright sunlight or when snow is on the ground he "can't hold his head up" owing to pain in his eyes. Heart and lungs normal; urine, acid 1022, no albumen. Oval piece of skin and some fibres of orbicularis muscle removed from right lower lid. This proved to be insufficient so that some weeks later another piece of skin was removed with a satisfactory result as five days later the lid edge was remaining well out. When visited 34 years later, in January, 1931, there was marked rolling in of right lower lid, Fig. 4, though the eyeball was quite quiet. No entropion of left lower lid. Has been working regularly on a steam line fishing boat and makes no complaint; eyelashes on all of the lids look normal, right palpebral fissure is not enlarged, plica semilunaris and caruncle are normal. Right lower lid is not thinned; R.V. $= < \frac{6}{60}$, with $\frac{-12 \text{ D. sph.}}{-2{\cdot}50 \text{ D. cylinder}}$ axis horizontal $= \frac{6}{18}$; L.V. $\frac{6}{6}$ fully with $+0{\cdot}50$ D. sph. Ophthalmoscopical examination: in right eye is a posterior staphyloma with complete choroidal atrophy, left fundus normal. Eyeballs of full size; no epicanthus, large fold of skin on both upper lids greater at outer part, no nodules felt in it, skin on forehead furrowed probably from attempts to keep up the lid skin fold, margin of upper lid is in good position, but especially on right side it becomes covered to a large extent by the skin fold, which requires to be raised by a finger for examination of fundus, considerable wrinkling of skin of lids. On right side the lashes on lower lid are completely hidden. Is married without issue.

The notes of the next case, first cousin of the above, are also from the hospital records. Wm. R. (III, 41), age 27 (June 30, 1896), fisherman. Diagnosis: R. entropion and corneal ulcer. In right eye at lower part of cornea is an ulcer, also old corneal opacity. About seven years ago patient noticed that lashes of his right lower lid were rubbing against the eyeball, causing the eye to become red and inflamed with much watering at times. Small ulcers appeared occasionally at the lower part of right eye where lashes irritated it. These usually soon disappeared. Seven weeks ago he noticed a fresh ulcer in right eye which as usual was red and watering and this time the ulcer did not go away as formerly. Three years ago there was an ulcer at the margin of his left cornea. The opacity spread over nearly the whole of the cornea but under treatment "has quite disappeared." His paternal grandfather (I, 3) had lower lids turned in (patient thinks he was born with it). Father is believed to have had affection of lids—one male first cousin (III, 51) had both lids and another

(III, 53) one lid turned in from birth. Patient as an infant had bronchitis and inflammation of the lungs. He has been quite healthy ever since, but never very strong. He looks healthy and well nourished. Heart and lungs normal; urine, acid 1022, no albumen, no deposit. An oval piece of skin was removed from right lower lid. The corneal ulcer soon healed and the lid margin kept in proper position. 14. 1. 31. He died unmarried.

Alexander R. (III, 51) born in 1871, fisherman. From an entry in the ophthalmic record book at the Aberdeen Royal Infirmary it was found that he was admitted as an in-patient on July 30, 1877, for "trichiasis." When seen on Jan. 20, 1931, he remembered being sick after the chloroform and having to return to hospital for removal of stitches from his lids. His lids were in good position, a scar could be detected on right lower lid and a smaller one was present on left lid; puncta, caruncle and plica semilunaris are normal; consistence of lower lids feels normal on palpation; no epiblepharon.

Of the next six cases we have the assurance of Dr MacHardy, who has worked among these people since 1892 and knows them intimately, that they (i—vi), now deceased, had entropion; additional reports regarding them given by the relatives are as follows:

(i) George R. (III, 39) had entropion of lower lids.

(ii) George A. (III, 57) married twice, no issue from either marriage, marked entropion.

(iii) George R. (II, 13), according to his niece Mrs W. (III, 47), had both lower lids turned in. In the notes of his son, Wm. R. (III, 41), the lids of his father were reported to be unaffected, but Dr MacHardy is certain that they were affected. It is possible that though he had entropion he experienced little inconvenience as was the case with David R. (III, 53), who had worked steadily for many years without discomfort with entropion of his right lower lid. Another possibility is that the entropion occurred periodically and was not observed by the son. No members of his sibship are alive.

(iv) John R. (II, 18), fisherman, according to his daughter, Mrs F. (III, 66), her husband, and mother-in-law (not shown in chart) had blinking eyes, but none of them could say there had been entropion. Vision became defective in advanced life.

(v) Mary A. (II, 16), according to her daughter, Mrs H. (III, 60), had both lower lids rolled in. She married her first cousin II, 19.

(vi) George R. (I, 3) married Mary A. (1, 4). Both lower lids were rolled in. His grandson's, Wm. R. (III, 41), notes state that his paternal grandfather's lower lids turned in and it was thought that he was born with the condition. A similiar note occurs in the records of David R. (III, 53, Plate I, Fig. 4), another grandchild.

Alexander R. (II, 15) died of tumour of the bowel at age of 54. His eyes were "the same as his brother John's" (II, 18). Dr MacHardy had seen him, but could not recollect in what way his eyes were affected. They certainly gave him trouble.

Mrs Robbie H. (I, 1). Her son Robbie H. (II, 1) said that his mother's eyes were always watering but he could not say whether the lids rolled in.

Jessie F. (III, 37), age 72 (1931). Right lower lid turned in chiefly at outer part, eye is blind, whole of cornea is opaque, eyeball is not shrunken, iris and pupil are visible. Ocular conditions attributed by her to an injury of right eye by a stone at age of five. No scars found suggesting an injury and her doctor suggests that she does not like to admit that the appearance of her eyes is an hereditary blemish. The opacity of the cornea may be explained by the lashes rubbing on the eyeball and causing inflammation. Eyelashes are normal; large skin fold on both upper lids, its lower edge passes from above inner canthus obliquely down and out and covers lower margin of upper lid for a considerable distance. Right palpebral fissure narrower than left; plica semilunaris and caruncle are normal, lower lids are not thinned; the entropion of right lower lid is not constantly present, but is obvious when she laughs; no epicanthus. Plate III, Fig. 5.

Isy R. (III, 42), age 61, sister of the above. Well-marked skin folds on upper lids, some loss of elasticity. Lids swell from time to time, swelling lasts for half a day, left lower lid has a fullness at its margin like the appearance seen in lids with spasmodic entropion. The fullness is not present on right lower lid, eyeballs normal externally, good vision. Her husband said that one of her lower lids sometimes rolled in and it was the left one. Previous information received from others was to the effect that her lower lids rolled in so that the lashes rubbed on her eyes.

George C. (IV, 50), age 54, fisherman, said that his eyes began to trouble him at the age of five. All of the lids became swollen, but the lower ones were worse than the upper. The swellings were intermittent. The eyes began to improve at age of 16; sometimes edges of lower lids rolled in. He spoke of the "family eyes" and on being questioned whether by that he meant a rolling in of the lower lid or a large fold of skin on upper lid he replied, "Both." He has now a small fold on upper lids. Skin of both upper and lower lids shows some wrinkling and loss of elasticity, colour of upper lid is reddish blue. Near lid margin, especially on left side, skin feels rather thick; no nodules felt; brow furrowed; eyes not raised. Plate III, Fig. 6.

Besides those cases with unusual appearances of the skin of the lids, that are associated with entropion, and which sometimes suggest blepharochalasis, there are a number of similar cases without entropion.

Mrs H. (III. 60), age 56, has had intermittent swellings of her lids. She cannot say definitely when they began, but they increased in size during menstruation. The skin of all the lids is much wrinkled, thin and deficient in elasticity, colour normal. No skin fold on upper lids. Thyroid not enlarged. No double lip. Has had good health, no digestive or nerve troubles, no migraine. Plate III, Fig. 7.

John H. (IV, 96), age 27, has a marked fold of skin on each upper lid. They overlap the upper lid margins at the outer and middle thirds. There is not much wrinkling of the skin which has its usual colour and consistence, no veins seen, no nodules felt; brow is furrowed, eyebrows lie low and straight, not arched. He is myopic, large crescent at lower part of margin of each optic disc. Plate III, Fig. 8.

Miss H. (IV, 97), age 25, marked fold of skin on each upper lid, especially the left one, which covers the outer two-thirds of the upper lid margin, skin is not wrinkled, no veins visible, colour not unusual, no nodules felt, consistence normal, no loss of elasticity. Lower lids show some furrowing. Formerly her lower lids swelled, lashes normal, refraction myopic; general health good, no migraine, no digestive or nerve disturbance. Thyroid not enlarged, no double lip. Plate IV, Fig. 9.

Mrs Elsie H. (III. 45), youngest in sibship of nine, age not recorded, but her sister III, 42, fourth youngest, is 61; marked skin folds on upper lids. The skin is thin and crinkled on both lower and upper lids; horizontal furrows on brow. Plate IV, Fig. 10.

Mr H. (IV, 63), one of her sons, has marked skin folds on upper lids.

James R. (V, 64), age 14. From eyebrows downwards nearly to lid margins is a swelling, skin at this part is reddish, and has lost some of its elasticity, no veins visible, no crinkling of the skin, nothing abnormal as regards its consistence, no nodules felt in the swellings, eyebrows not raised, no entropion, some furrows on lower lids, there is blepharitis (the only case in the pedigree in which this was observed). Eyes affected since he was three years old. Swelling has not been intermittent. Health good, no digestive or nerve affection. Plate IV, Fig. 11.

George R. (V, 65), age 13. Eyes became affected when about two years of age. There was intermittent swelling of both lower and upper lids. Has had good health. Skin fold of upper lid is greater on left side than on right. The condition is said to be less marked now than formerly. Skin of upper and lower lids shows numerous horizontal lines, colour and consistence of skin normal, no loss of elasticity. Neither he nor his brother James has double lip or enlarged thyroid.

Master M. (V, 71), age 7. Well-marked swelling like folds on both upper lids. Plate IV, Fig. 12.

William P. (IV, 8), age 17. First born of five. Marked swelling below eyebrow on each side. Plate II, Fig. 13. It obscures outer two-thirds of upper lid margin. The illustrations of Shoemaker's case of bilateral enlargement of the lachrymal glands and of Wagenmann's case of ptosis adiposa resemble the appearances presented by IV, 8 and V, 64. George P. (III, 5), father of IV, 8, has much wrinkling of skin of lids and a fold of low degree on upper lids. Has five children. First-born age 17.

Maggie S. (IV, 45) has obliquely lying skin folds on upper lids. Her children, Mary S. (V, 33) and Jacob S. (V, 34), have marked fullness under the eyebrows. Ages about 18 and 17 years respectively.

Jessie M. (Mrs S. IV, 46), age 43, on upper lids lax, redundant skin with very marked loss of elasticity.

Georgina Mac. (IV, 49), age 31, bulging of tissue below eyebrow. No swelling of lower lids.

William Mac. (V, 60), child of IV, 49. Right side, swelling below eyebrow is greater than that on left side. It presses on the lashes when his head is held erect and he looks at an object on the same level as his eyes.

Alexander Mac. (V, 61), child of IV, 49, fold of skin on right upper lid which touches lid edge at its middle portion.

Annie H. (I, 1) was reported by her son II, 1 to have always had watering eyes. Whether the lids rolled in could not be said. There was an indefinite history that I, 2 had been affected in the same way.

When examining the upper eyelids of the relatives it became evident that a standard was required to serve as a guide to what should be regarded as a normal and what an abnormal appearance of the upper eyelid. It was decided to classify as normal those eyelids whose margins were visible and not obscured at any part by swelling or bulging of the skin or by a skin fold. To form some estimate of the proportion of eyelids with such abnormal appearances in the general population 502 people were examined. These were taken almost exclusively from the general wards of the Aberdeen Royal Infirmary and the Royal Aberdeen Hospital for Sick Children. In such an examination there are several pitfalls. Should the examinee look even slightly upwards the margin of the lid, previously visible, may in some cases become obscured. Similarly, when the examinee is fixing an object when the visual axes are parallel with the ground a very slight depression of his head may cause some of the lid margin to be hidden. Very important is the position of the head, which must be held erect during the examination. Another important matter when examining doubtful cases is the position of the examiner's eye, which must be on the same level as the examinee's eyes and the object he is fixing. In some cases in which part of the lid margin becomes temporarily obscured frowning is the cause. Some children exposed to bright sunshine had the margin of the lid hidden, but when brought into the shade the margin became visible. By raising their eyebrows some individuals can readily expose the upper lid margin previously covered by a fold of skin.

The 502 control cases of all ages included 20 persons, 12 males and 8 females, with some obscuration of the margin of the upper lid on one or both sides, namely 3·98 per cent. In dividing the cases into two age groups there are 254 people aged 20 years and upwards and 248 people aged from infancy to 19 years. In the former group are 10 abnormal cases, that is the upper lids had an abnormal appearance, namely 3·93 per cent., in the latter group are 10 abnormal cases, namely 4·43 per cent. Of the 1004 upper eyelids examined 34 were abnormal, i.e. 3·38 per cent.; 6 of the cases were unilateral. In the group of older people were 16 abnormal upper lids, 3·14 per cent., and in the group of younger people were 18 abnormal upper lids, 3·62 per cent. The ages of those with part of the margin of the upper lid hidden by the skin of the lid at time of examination were: In first decennium, 7, of whom 4 males, 3 females; second decennium, 3, all males; third decennium, 2 females; fourth decennium, 3, of whom 1 male, 2 females; sixth decennium, 2, of whom 1 male, 1 female; seventh decennium, 1 male; eighth decennium, 2 males.

The pedigree chart shows 13 individuals with entropion, 3 females and 10 males. These are distributed in four consecutive generations and invariably the condition passes directly from parent to child, usually from father to son or daughter. In one

instance only has an affected female had an affected child; II, 16, a female, had a son, III, 57, with entropion. Generally the entropion was noticed in early life and in some cases it was believed to have been present from birth. Owing to the insignificant disturbance caused by entropion in early life it is probable that in a number of cases the condition had been present for a longer period than the history indicates. Aubineau found symmetrical entropion in an infant of 10 months, that manifested no pain, and appeared little disturbed, and the rubbing of the lashes produced only a slight injection of the conjunctiva below the cornea. Again, in adult life entropion may be present for many years without causing much discomfort or interfering seriously with work as was seen in the case of D. R., III, 53, a fisherman. The entropion in some of the cases was not constantly present. In all cases it occurred in the lower lid. In some it was bilateral, in others unilateral.

Since an abundance of extensible skin is necessary for the development of spasmodic entropion found commonly in elderly people, it is presumably the same in cases occurring in early life, such as the examples seen in this pedigree. A number of cases occur in the pedigree with loose skin on the lid. Several cases had relaxed, wrinkled skin on the lower lid after intermittent swelling. If the skin of the lower lid becomes relaxed two factors would tend to cause entropion, namely strong contraction of the palpebral portion of the orbicularis muscle and an insufficient supply of orbital fat. The first factor would operate, especially in the men when at work as fishermen, from exposure to bad weather, spray, sun, and reflection from the sea. This factor may be of importance in this pedigree. Insufficient orbital fat by causing the eyeballs to lie far back in the orbit diminishes the support for the lids. In my experience these fisher folk are not generally of robust appearance and many of them are small. I am informed that sometimes they obtain insufficient nourishment, also that they are injudicious eaters, consuming many cakes and sweatmeats when they can afford to buy them. Whether these habits are of any importance in the causation of this condition I cannot say. No evidence has been obtained that bears upon the matter.

In the pedigree occur several unusual appearances of the skin of the eyelids. These changes are found in those without entropion as well as in those with entropion. Three forms are distinguishable. The first form occurs in the upper lids and is bilateral. There is obvious swelling extending from the supra-orbital margin downwards so as to obscure the tarsal portion of the eyelid and part of the lid margin, often the middle and outer thirds. The swelling appears as a fullness or bulging and not as a flap. On palpation no nodules are felt in the swelling. The skin has a normal appearance or else is redder than usual, few if any veins are distinguishable. Two examples of this first form, which may be the earliest stage of blepharochalasis or epiblepharon, are seen in Plate IV, Fig. 11 and Plate II, Fig. 13. The recognition of the first form, unless it is well marked, may be difficult, or impossible, owing to its close resemblance to apparently normal conditions found in a proportion of the general population. Amongst 502 individuals whose ages ranged from infancy to 81 years, in several a similar bulging of the skin extended so low

as to touch the lashes and hide part of the lid margin, when the individual, with head held erect, looked straight forward at an object on the same level as the eyes. But no such difficulty arises in such marked examples as V, 64 (Fig. 11) and IV, 8 (Fig. 13). The second form is a fold of thin skin with or without wrinkling, that has lost elasticity, and is situated at the usual site of the normal skin fold of the upper lid. Its lower margin passes obliquely from above and inwards in a direction downwards and outwards so as to cover the edge of the lid at its outer and sometimes middle third. See Plate IV, Fig. 10. This form might be termed ptosis atonica or epiblepharon and in some cases possibly blepharochalasis. In the sister and brother, IV, 97 and IV, 96, with marked folds on their upper lids, it could not be determined whether they had been preceded by swelling of the lids. Their mother, III, 60, who had noticed nothing peculiar about her children's eyelids, had noticed no swelling. The third form occurs in both upper and lower lids and resembles the wrinkled skin of old age. The skin is thin, shows loss of elasticity and is wrinkled. See Plate III, Fig. 7, a woman who had repeated attacks of swelling of both upper and lower lids. As these three clinical forms occur amongst those who are all consanguineous relatives it is just possible that they are different stages of the same affection or at any rate that one form may develop into one of the other two forms. There is no proof, however, that they do so, and to decide the matter definitely, it would be necessary to observe individuals with the first form for prolonged periods. That the first form becomes the second or third forms is suggested by the earlier age of its occurrence. The average age, when seen, of six of the seven cases in the first group in which the age is recorded is 14·33 years. The seventh case was 31. The average age of cases in the second group was 47·80 years and in the third group 49·00 years. These are the ages when the cases were seen and not when the appearances were first manifested. The wrinkled skin (third form) which occurs in the lower lid obviously cannot result from the swelling of the upper lid (first form). The wrinkling of both upper and lower lids probably resulted from intermittent swellings of these lids. (Cases III, 42, IV, 50 and V, 65.)

Benedict says that blepharochalasis has an intumescent stage or stage of oedema. After several attacks of swelling one of two things may happen. (1) Swelling becomes constant with bagginess of the skin of the lid so that loose folds hang down over the margin, giving the appearance of water-filled bags, with the skin altered slightly in colour, very thin, and slightly folded or wrinkled; or (2) the swelling disappears entirely and the skin becomes reddish brown and wrinkled and is thrown into horizontal folds. The stage of wrinkling is the end stage of the disease.

Our chart shows 21 cases with abnormal conditions of the skin of the lid; 12 are males and 9 females; 4 have entropion and 17 no entropion. These figures almost certainly underestimate the number of cases present in the pedigree, since doubtless a number of individuals in generations I and II had similar changes in the skin of the lids, and in generations III to V cases may be present amongst those who were not examined, and also in some members in generations IV and V the condition may not yet have appeared.

The question arises, what is the significance, if any, of the presence in the same pedigree of both entropion and these changes in the skin of the lids? A glance at the small skeleton chart below (Fig. 14) shows at once the relationship of one to the other. Besides the occurrence of each of these conditions separately there are at any rate four cases in which they occur together in the same individual. If it be accepted that all cases of spasmodic entropion require for their occurrence relaxation and redundance of skin of the lid then all of the entropion cases in this pedigree must be classed with the cases grouped under "changes in the skin of the lids," which would now number 30. Twenty-four of these, in four generations, show direct inheritance from an affected

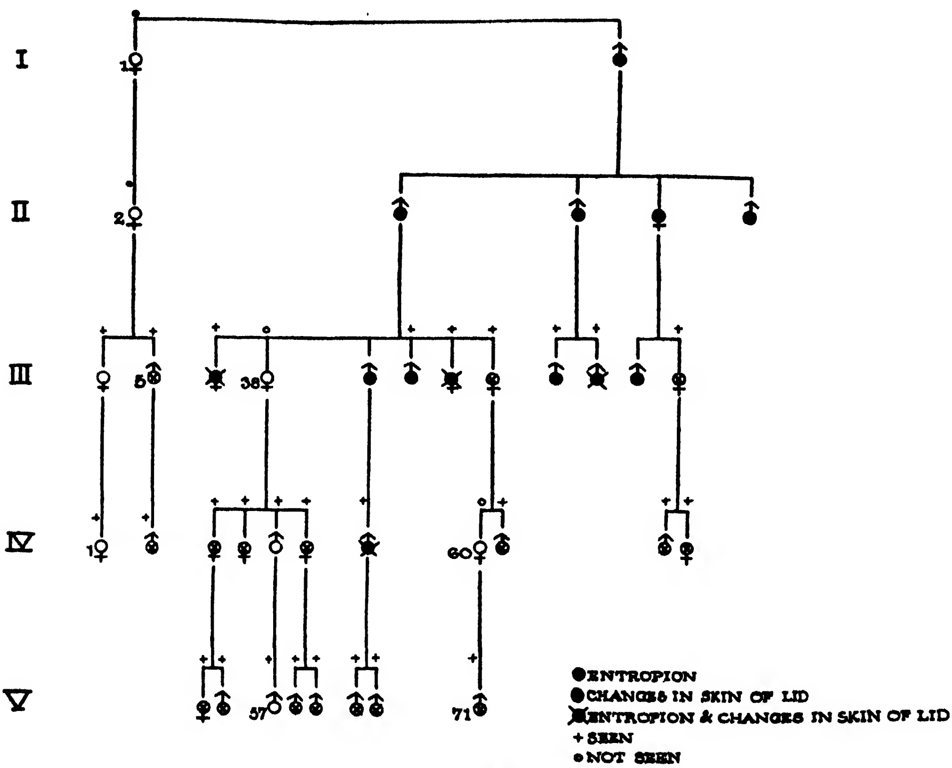


Fig. 14.

parent, and of the remaining five cases it cannot be said that the changes in the skin of the lid were not directly inherited from their mothers, II, 2, III, 38 and IV, 60, for they were not examined. It seems reasonable to suppose from the close association of the changes in the skin of the eyelids and entropion that a factor is inherited which in some way unknown is responsible for the presence of the former and that the entropion is only a secondary manifestation. It may be thought that there are no good grounds for the supposition for most of the changes recorded in the skin were in the upper lids while the entropion was invariably in the lower lids. In a proportion of the cases, however, definite changes were seen in the skin of the

lower lids and also quite possibly changes sufficient to permit the occurrence of entropion may take place in the skin of the lower lid without showing marked clinical manifestations. I am informed that fisher families in the same village that do not belong to this pedigree are not subjects of hereditary entropion.

Summary.

Reference is made to cases in the literature which have a bearing on various forms of change in the skin of the lids and on congenital entropion. Two cases not belonging to the pedigree are reported of epiblepharon or blepharochalasis with double lip.

A pedigree with chart and photographs is shown which contains cases of spasmodic entropion appearing at unusually early ages and cases with changes in the skin of the lids. Three varieties of the latter are recognised. First a swelling in the form of fullness or bulging of the skin of the upper lid, second a skin fold of the upper lid, and third a crinkling of the skin of both upper and lower lids. The entropion is invariably transmitted from parent to child, usually from father to son or daughter, and similarly the above described changes in the skin of the lids, which occur in cases with or without entropion, show continuous descent*.

Attention is drawn to the importance when examining folds of skin, or swellings, of the upper lid in relation to the lid margin of recognising the alterations in relative position that occur when the head is not erect, when the position of the eyes are altered, when frowning occurs or when the eyebrows are raised and when the examiner's position is altered.

Owing to the close association in the same pedigree and sometimes in the same cases of changes in the skin of the lids and entropion it is suggested that there is an hereditary factor that in some unknown way is responsible for the changed condition of the skin of the lids and that the occurrence of entropion is only secondary.

My thanks are due to Dr T. MacHardy for much help in tracing out the pedigree and for information regarding deceased affected members, to Dr J. W. Macrae for showing me so many descendants of III, 38, I, 1 and I, 2, and to Dr William Brown for the photographs shown in Plates I—IV, Figs. 2 to 12.

REFERENCES.

ACCARDI. Blepharochalasis. *Bollettino di Oculistica*, No. 6. Firenze, 1925. Ref. *Annales d'Oculistique*, T. CLXV. p. 224. Paris, 1928.

v. AMMON, F. A. *Klinische Darstellungen der angeborenen Krankheiten des Auges und der Augenlider*. 6, Tab. II. Fig. XV. Berlin, 1841.

ARLT, C. F. *Graefe-Saemisch Handbuch der ges. Augenheilkunde*, Bd. III. S. 454. Leipzig, 1874.

ASCHER, KARL W. Blepharochalasis mit Struma und Doppellippe. *Klinische Monatsblätter f. Augenheilkunde*, Bd. LXV. ii. S. 86. Stuttgart, 1920.

——. Das Syndrom Blepharochalasis, Struma und Doppellippe. *Klinische Wochenschrift*, Bd. I. S. 2287 (Nov. 11). Berlin, 1922.

AUBINEAU, E. Les "Entropions congénitaux." *Annales d'Oculistique*, T. CLXV. pp. 161—169. Paris, 1928.

* [An exception to this rule appears in V, 71, who inherits from his grandmother III, 45, through a normal mother, IV, 60. ED.]

BACH, L. Ueber symmetrische Lipomatosie der Oberlider (Blepharochalasis?). *Archiv f. Augenheilkunde*, Bd. LIV. S. 73. Wiesbaden, 1906.

BENEDICT, WILLIAM L. Blepharochalasis. Report of three cases. (Sect. of Ophthalmology, Mayo Clinic). *Journal of American Medical Association*, Nov. 1926, p. 1735. Chicago, 1926.

BERRY, GEORGE A. *Diseases of the Eye*, 2nd edition, p. 56. London, 1893.

DENIG, R. Beitrag zur Aetiologie der angeb. Trichiasis. 71. *Versamml. der deutschen Naturf. und Ärzte in München. Abteil f. Augenheilkunde*, 1899. Ref. Nagel's *Jahresbericht der Ophthalmologie*, Bd. XXX. S. 335. Tübingen, 1899.

DIMMER, F. Epicanthus und Entropium. *Klinische Monatsblätter f. Augenheilkunde*, Bd. XXIII. S. 308. Stuttgart, 1885.

EIGEL, WALTER. Blepharochalasis und Doppellippe ein thyreotoxisches Oedem? *Deutsche medizinische Wochenschrift*, Bd. LI. S. 1947 (Nov. 20). Leipzig, 1925.

FEHR. Ein Fall von Lidhaut-Erschlaffung, sog. Blepharochalasis. *Centralblatt f. praktische Augenheilkunde*, Bd. XXII. S. 74. Leipzig, 1898.

FRANCESCHETTI, A. *Kurzes Handbuch der Ophthalmologie* (herausgegeben von F. Schieck und A. Brückner), Bd. I. Berlin, 1930.

FRIEDENWALD, J. S. Further note on Blepharochalasis. *Archives of Ophthalmology*, Vol. LII. p. 367. New York, 1923.

FUCHS, E. Ueber Blepharochalasis (Erschlaffung der Lidhaut). *Wiener klinische Wochenschrift*, Bd. IX. S. 109. Wien, 1896.

GINESTOUS. Un Cas de Blépharochalasis. *Gaz. hebd. des Scienc. méd. de Bordeaux*, 27 juillet, 1913. Ref. Nagel's *Jahresbericht der Ophthalmologie*, Bd. XLIV. S. 359. Tübingen, 1915.

GOMEZ, MARQUEZ. Congénital Entropion. *Archivos de Oftalmología Hispano-Americanos*, Vol. XXX. p. 187. Madrid, 1930. Ref. *Archives of Ophthalmology*, Vol. IV. p. 648. New York, 1931.

GRAF. Oertliche erbliche Erschlaffung der Haut. *Wochenschrift f. d. ges. Heilkunde*, No. 15 (April 9). Berlin, 1836.

GUIBERT. Un Cas d'Entropion congénital double. Guérison. *Archives d'Ophtalmologie*, T. XII. p. 101. Paris, 1892.

HARLAN, GEORGE C. *Transactions, American Ophthalmological Society*, Vol. VII. p. 418. 31st Annual Meeting, New London, 1895.

HECKEL, EDWARD B. Blepharochalasis with Ptosis; Report of a Case. *American Journal of Ophthalmology*, Vol. IV. p. 273. Chicago, 1921.

HESSBERG, RICHARD. Ueber ein angeborenes familiäres Entropion beider Unterlider. *Klinische Monatsblätter f. Augenheilkunde*, Bd. LXVIII. i. S. 120—124. Stuttgart, 1922.

HIMLY, K. *Die Krankheiten und Missbildungen des menschlichen Auges*. Bd. I. S. 121—122. Berlin, 1843.

JENISON, NANCY. Blepharochalasis. *New York Medical Journal*, Vol. CII. pp. 555—556. New York, 1915.

LAFON, CH. et VILLEMONT. Blépharochalasis héréditaire avec Dacryadénoptose. *Archives d'Ophtalmologie*, T. XXVI. p. 639. Paris, 1906.

LEBLOND, ÉTTIENNE. Étiologie de l'entropion congénital. *Archives d'Ophtalmologie*, T. XXVII. pp. 782—787. 1907.

LODATO, G. Blépharochalasis. *Archivio di Ottalmologia*, Vol. XI. Napoli. Ref. *Archives d'Ophtalmologie*, T. XXIV. p. 770. Paris, 1904.

LOESER, L. Ueber Blepharochalasis und ihre Beziehung zu verwandten Krankheitsbildern nebst Mittheilung eines Falles von Blepharochalasis mit Spontan-Luxation der Thränendrüse. *Archiv f. Augenheilkunde*, Bd. LXI. S. 252. Wiesbaden, 1908.

MACKENZIE, W. *Diseases of the Eye*. 4th edition, p. 187. London, 1854.

MEGURO, T. Ein Fall von Fetthernien der beiden oberen Augenlider. *Acta Soc. Op. Jap.* Vol. XXXIV. p. 13, 1930. Ref. *Klinische Monatsblätter f. Augenheilkunde*, Bd. LXXXVI. i. S. 843. Stuttgart, 1931.

MIGLIETTA, F. Su la Blefarocalasi (Contributo clinico ed anatomo-patologico). (Clin. Oculist. Univ. Firenze.) *Bollettino di Oculistica*, Vol. IX. pp. 1190—1214. Firenze, 1930.

MÜLLER, H. K. Kongenitales Entropium durch Epiblepharon. *Klinische Monatsblätter f. Augenheilkunde*, Bd. LXXXVII. S. 184. Stuttgart, 1931.

POLSTOOCHOW. Entropion congénital familial de la Paupière inférieure. *Vestnik Oftalmologii*, pp. 820—830. 1914. Ref. *Archives d'Ophtalmologie*, T. XXXV. p. 125. Paris, 1916—17.

ROHMER. De l'Angio-mégalie symétrique des Paupières supérieures. *Archives d'Ophtalmologie*, T. XX. p. 407. Paris, 1900.

ROSENSTEIN. *Centralblatt f. praktische Augenheilkunde*, Bd. XXVI. S. 233. Leipzig, 1902.

SCHMIDT-RIMPLER, H. Fett-Hernien der oberen Augenlider. *Centralblatt f. praktische Augenheilkunde*, Bd. XXIII. S. 297. Leipzig, 1899.

SCHORR, ILSE. Ueber Entropium spasticum des Oberlides. *Klinische Monatsblätter f. Augenheilkunde*, Bd. LXXVI. i. S. 373. Stuttgart, 1926.

SCRINI. Un Cas de Blépharochalasis (Ptosis atonique. Dermatolysie palpébrale). *Archives d'Ophtalmologie*, T. XXVI. pp. 440—446. Paris, 1906.

SHOEMAKER, W. T. A Case of Bilateral Enlargement of the Lachrymal Glands. *Annals of Ophthalmology*, Vol. XIII. p. 513. St. Louis, 1904.

SICHEL, J. *Annales d'Oculistique*, T. XII. p. 188. Paris, 1844.

STEPHENSON, SYDNEY. Note upon a form of Congenital Trichiasis. *Transactions, Ophthalmological Society of U. K.*, Vol. XIV. p. 13. London, 1894.

STIEREN, EDWARD. Blepharochalasis. Report of two Cases. *Annals of Ophthalmology*, Vol. XXIII. p. 625. St Louis, 1914.

SZIKLAI, PAUL. Zur Frage des Entropium congenitum. *Zeitschrift f. Augenheilkunde*, Bd. XXXVIII. S. 103. Berlin, 1917.

TERSON, A. Dermatolysie palpébrale. *Archives d'Ophtalmologie*, T. XXIV. p. 346. Paris, 1904.

VERHOEFF, F. H. and FRIEDENWALD, J. S. Blepharochalasis. *Archives of Ophthalmology*, Vol. LI. p. 554. New York, 1922.

WAGENMANN, A. Ein Fall von doppelseitiger echter Ptosis adiposa bei einem 16 jährigen jungen Mann. 34. *Vers. der ophth. Gesellschaft*, S. 274. Heidelberg, 1907.

WEIDLER, W. B. Blepharochalasis. *Journal of American Medical Association*, Vol. LXI. pp. 1128—1132, Sept. 27. Chicago, 1913.

WEINSTEIN, A. Über zwei eigenartige Formen des Herabhängens der Haut der Oberlider: Ptosis adiposa mit spontaner Senkung der Tränendrüse. *Klinische Monatsblätter f. Augenheilkunde*, Bd. XLVII. ii. S. 190. Stuttgart, 1909.

WEVE. Een Geval van Syndrom Ascher. *Nederl. oogheelkundig Jaarboek*, 1921. Ref. Ascher, *Klinische Wochenschrift*, Bd. I. S. 2287 (Nov. 11). Berlin, 1922.

WIRTHS, M. Beiderseitige Lidgeschwulst, kombiniert mit Geschwulstbildung der Oberlippe. *Zeitschrift f. Augenheilkunde*, Bd. XLIV. S. 176. Berlin, 1920.

YAN CHOW. Ueber Entropium der Neugeborenen. *Klinische Monatsblätter f. Augenheilkunde*, Bd. LXXV. ii. S. 162. Stuttgart, 1925.

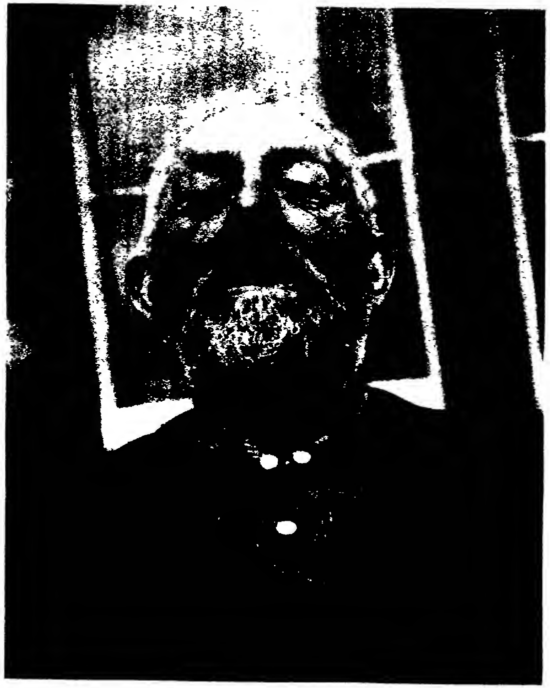


Fig. 1.

Fig. 3.

Fig. 4. III. 53.

Fig. 2.

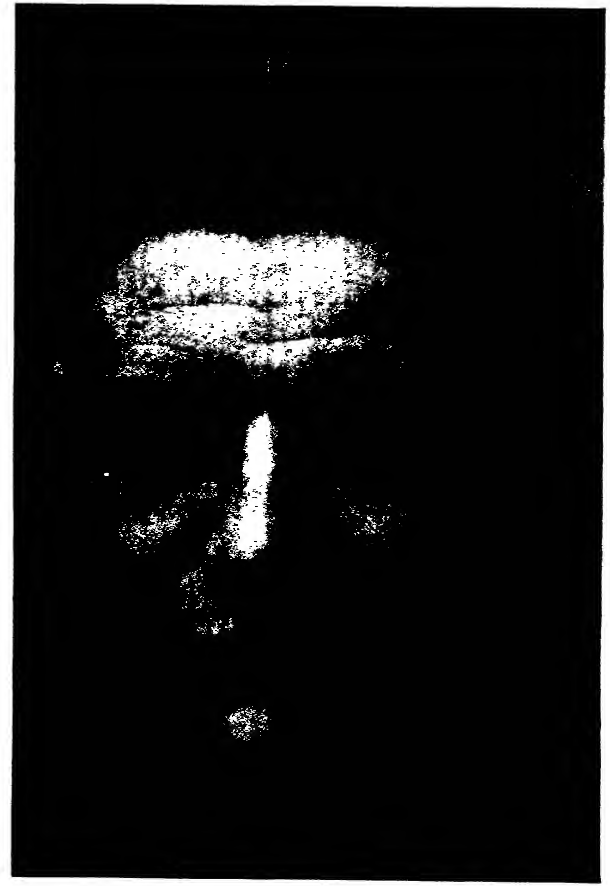


Fig. 13. IV, 8.

Fig. 5. III, 37.

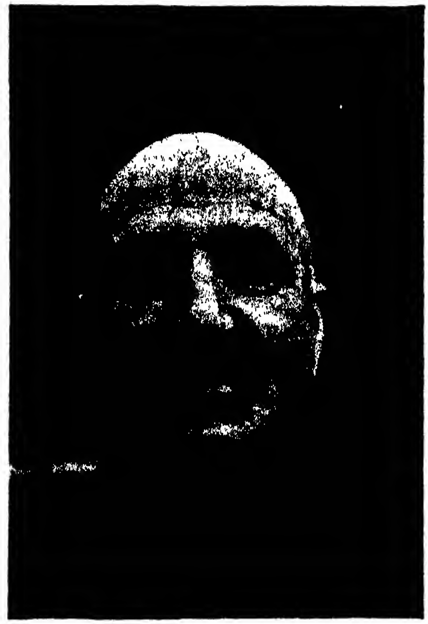


Fig. 6. IV, 50.

Fig. 7. III, 60.

Fig. 8. IV, 96.

Fig. 9. IV, 97.

Fig. 10. III, 45.

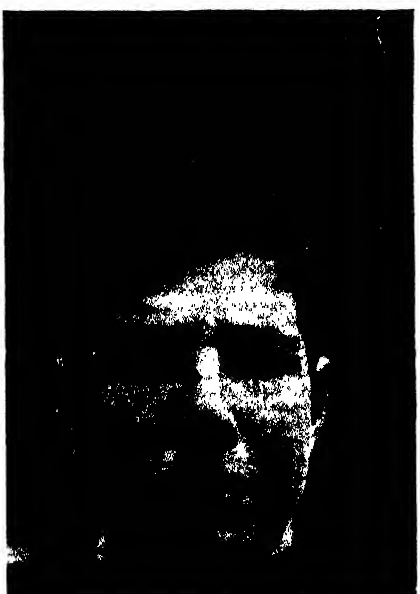


Fig. 11. V, 64.

Fig. 12. V, 71.

THE GEOMETRIC PROPERTIES OF MICROSCOPIC CONFIGURATIONS. I. GENERAL ASPECTS OF PROJECTOMETRY.

By WILLIAM R. THOMPSON.

(From the Department of Pathology, Yale University.)

For several years attention has been directed in this laboratory to a study of the geometric properties of configurations too small to be studied conveniently with the eye directly. The procedure has involved the projection of images of plane sections of the configurations which are subjected to measurement. For such procedure the name *Projectometry* is suggested.

As an example of an application of Projectometry may be cited the estimation of the volume of an island of Langerhans in a pancreas. The method employed is similar to that used in the estimation of the volume of the hull of a ship from cross-sectional plans to scale. The areas of parallel plane sections at definite intervals along a given perpendicular axis are estimated by means of measuring the areas of their magnified images with a planimeter. With such an area as ordinate and position of the plane of section relative to the given axis as abscissa a point is located on a co-ordinate plane in each instance. If the cross-sectional area of the given configuration were so determined at all points along the given axis and the appropriate scale used in the co-ordinate system, then the area bounded by the curve so obtained and the axis of abscissae would be numerically equal to the volume of the configuration. In practice, however, we employ a trend line as an approximation to the above mentioned curve; the simplest device, technically, being what is called a fitted smooth curve, a curve drawn by inspection of plotted points as free as practicable from personal prejudice. Errors thus introduced may be estimated roughly in most cases in practice by sensibly independent attempts to perform the same operation. An error of this sort is ultimately involved in all graphic methods but may usually be restricted to a negligible magnitude. If convenient, any other scale of plotting may be adopted, replacing the above ordinates and abscissae by h_1 and h_2 times their value, respectively, and dividing the resultant area by $h_1 . h_2$ to obtain the volume mentioned above. The result is invariant with translation of the axis of ordinates, which may be employed, therefore, whenever convenient. Obviously, areas measured by planimeter must be converted to expressions in terms of the appropriate units; e.g., if the area under the trend line is so measured (called the *Graph Area*) the direct result of measurement is divided by a similar measurement of the area of a unit square of the co-ordinate plane, or in either or both instances the single

measurement may be replaced by the mean of several independent observations. In any given experience under discussion it is preferable that a uniform treatment be employed.

In the following discourse the name *islet* will be used to designate a domain composed of a connected region in three dimensional space, possessing a finite volume, and its frontier. Obviously, the above-mentioned method may be applied to the estimation of the volume of such a configuration, provided that traces of images to scale can be made of sections of the islet intersecting a given axis perpendicularly at given intervals; or, at least, at intervals small enough to yield satisfactory approximation in the graph.

Now, we may have under consideration a domain having the attributes of an *islet* as defined above, but of primary interest in a given discourse with respect to the islet or islets of a specific kind that it may contain. Such a domain will be called the *organ* and the islets of the specified kind which it contains will be called its islets in such discourse. In such systems (e.g. the pancreas and its islands of Langerhans) many interesting properties may be studied *, it being assumed that the islets are discrete and finite in number as in the case cited, such as the arithmetic mean or other forms of average of the islet volumes, measures of dispersion, the distribution, aggregate number and total volume of the islets of the organ, volume of the organ and various ratios of these such as the ratio of total islet volume to organ volume and of aggregate number of islets to organ volume. Of course, a method of counting all the islets and measuring the volume of each might be devised; but the time and effort involved in such an undertaking might be prohibitive. In such cases the usual resort is to estimate certain of these quantities by means of similar data obtained from a random sample or samples of the whole, the function of these samples so estimated being chosen so that the appropriately weighted mean of values so obtained from mutually exclusive random samples converges to the value of the same function of the whole universe of discourse as the sum of the included samples approaches the universe as a limit. Now, by definition, a random sample of a universe of a finite number of things is a sample such that the *à priori* probability of being selected is the same for each individual in the universe. However, it is possible to obtain nearly if not quite as satisfactory estimates of functions of a universe from samples not necessarily having this property, provided that the various *à priori* probabilities of selection of the individuals actually selected in a given experience are known or at least their relative values, that the selection is uninfluenced by prejudice as to the final result, and that in the selection there is no individual in the universe whose probability of selection is zero. We may resort in such cases to a procedure of which the following is an illustration.

Let U be a universe of n individuals, $\{A_i\}$, numbered from one to n; and let

* Since the present article was written, the author's attention has been drawn to two reports by Wicksell (Wicksell, S. D., "The Corpuscle Problem," *Biometrika*, Vol. XVII. (1925), pp. 84—99, and Vol. XVIII. (1926), pp. 151—172), which deal, respectively, in a comprehensive manner with the special cases wherein the shape of the islets (or corpuscles) is spherical or elliptical.

x_i be a real number* corresponding to A_i, for $i=1, \ldots\ldots, n$. Then we define m, the arithmetic mean of the quantities $\{x_i\}$, by

$$m = m_{(x)} = \frac{\sum_{i=1}^{i=n} x_i}{n} \qquad \text{.................................(1).}$$

Now, let us replace the ordinary notion of sampling by the following notion of set selection.

Let individuals be selected from U, one at a turn, noting the result of each selection and then returning the individual to the same situation in regard to probability of selection at any turn as existed initially; i.e. let the probability of selecting any given individual of U be the same at each turn.

Let p_i, a constant, be the *à priori* probability of selecting the individual designated by A_i at any turn; and let f_i be the number of times A_i has been selected in a set, S, of r turns; where r may be any positive rational integer and $i=1, \ldots\ldots, n$.

Then, according to previous specification, we require that

$$p_i \neq 0 \text{ for every } i, \text{ and } \sum_{i=1}^{i=n} p_i = 1 \qquad \text{....................(2).}$$

Now in a set, S, of r turns, obviously, it may happen that for some value or values of i, $f_i=0$. Indeed, this is a necessity if $r<n$.

In any case, let us adopt the following definition for $M_{(x)}$, for which we shall use M as an abbreviation.

$$M = M_{(x)} = \frac{\sum_{i=1}^{i=n} \left[f_i \dfrac{x_i}{p_i} \right]}{\sum_{i=1}^{i=n} \left[\dfrac{f_i}{p_i} \right]} \qquad \text{.(3).}$$

Now, as we have defined p_i as the *à priori* probability of selection of A_i at any given turn, we have

$$\lim_{r=\infty} \left[\frac{f_i}{r} \right] = p_i \qquad \text{.(4).}$$

Indeed, the assumption of the relation (4) suffices as a definition of p_i. Accordingly, by dividing the numerator and denominator of the right member of (3) by r, we obtain by (1) and (3)

$$\lim_{r=\infty} M = m \qquad \text{...................................(5).}$$

This last property is necessary to any satisfactory estimate of the mean m, by sampling; but, as is the case with the mean of a random sample of U, here also we have no absolute knowledge of the error of the approximation. In either case we do not know how good is the convergence; but in either case we have the same resource, namely, to contrast several independent estimates (customarily not less than four), each being obtained according to the same technical plan.

* Immediate extension to any system of magnitudes of uniform dimension (e.g. mass, volume, length, etc.) may be made, obviously.

It is of interest to note another property of M in relation to the same function of mutually exclusive subsets of the given selection experience. Let there be a dichotomous division of S into two subsets which we shall call S' and S''. Let us provide furthermore

$$\begin{array}{rl} \text{that } M',\ f_i' \text{ and } r' & \text{be related to } S' \\ \text{and that } M'',\ f_i'' \text{ and } r'' & \text{be related to } S'' \\ \text{as } M,\ f_i \text{ and } r & \text{are related to } S. \end{array}$$

Then, obviously,

$$f_i' + f_i'' = f_i \text{ and } r' + r'' = r \qquad (6).$$

Now, let

$$a = \sum_{i=1}^{i=n} f_i \frac{x_i}{p_i} \quad \text{and} \quad b = \sum_{i=1}^{i=n} \frac{f_i}{p_i} \qquad (7);$$

and let

$$\begin{array}{rl} a' \text{ and } b' \text{ be} & \text{related to } S' \\ \text{and } a'' \text{ and } b'' \text{ be} & \text{related to } S'' \\ \text{as } a \text{ and } b \text{ are} & \text{related to } S \end{array} \qquad (8).$$

Then, by (3) we obtain

$$M = \frac{a}{b}, \quad M' = \frac{a'}{b'} \quad \text{and } M'' = \frac{a''}{b''} \qquad (9),$$

and, obviously, $a' + a'' = a$ and $b' + b'' = b$, whence we have

$$M = \frac{b'M' + b''M''}{b' + b''} \qquad (10),$$

which is the relation obtained if S' and S'' were merely subsets of S defined as a set of $b' + b''$ numbers (or magnitudes), S' and S'' being mutually exclusive and containing, respectively, b' and b'' of the elements of S, and M, M' and M'' being, respectively, the arithmetic means of the elements of S, S' and S''.

Now, there is another extremely valuable property of M which is at once obvious when we note that in (3) wherever $f_i = 0$ we need not know the value of p_i as it has been provided that $p_i \neq 0$. This means that the *à priori* probability of selection need be known only for those individuals actually selected, and indeed only their relative values need be known in order to calculate $M_{(x)}$ from the observed values of x_i and f_i in the experience, S.

The above is by no means the only case wherein the failure of the ordinary method of random sampling to be readily applicable may be offset by an application of a knowledge of the *à priori* probabilities of selection of individuals. The function, M, defined in (3), is of like value in the case where more than one individual is selected at a turn, provided that the appropriate relative *à priori* probabilities of selection are used.

We turn, now, to another feature of the study of the properties of islets in relation to an including organ in the general sense adopted above, namely, the estimation of *population density* and location of islets. Such a subject would be extremely difficult if not unmanageable were discourse not restricted to some arbitrarily designated principal point of an islet. Obviously, such a point should be defined uniquely.

For the present, at least, let us assign this rôle to the centroid of the islet. This point is determined uniquely and depends upon geometric properties only. In the case of an islet of uniform density, of course, the centre of gravity and centroid coincide.

Accordingly, we shall define the population density of islets in a region by the ratio of the number of islet centroids lying in the region to the volume of the region. In general, where we may be dealing with a domain consisting of a region with part or the whole of its frontier, we shall say that an islet *belongs* to the domain if the domain contains the islet's centroid; and let ϕ be the population density defined as equal to the number of islets belonging to the domain under discussion divided by the domain volume (provided, of course, that the domain has a volume). It may be convenient to subdivide an organ into a set of mutually exclusive domains of this sort (having the whole organ equal to their sum), and in practice we may deal with the regional part of such domains, which have the same volume, and the manner of their choice for study can be such that the probability, *à priori*, of any islet centroids lying on the frontiers of the regions is zero (the actual incidence being neglected, accordingly).

Furthermore, in practice, we often resort to the estimation of a function by means of another function to which it bears a known or perhaps a tentatively assumed relation. In the application of the foregoing to the estimation of islands of Langerhans per unit volume at least two substitutes may be utilised in place of centroid location.

Consider the organ to be cut by a set of parallel planes at intervals of 10 micra, and let us agree to accept data obtained from a given section (domain equal to the part of the organ lying between two such successive planes) as indicative of the mean value of a given function (such as ϕ, the population density) in a given region; i.e. to regard the section as a fair sample of a domain including it. Let T be the total tissue volume of the section and N be the number of islets belonging to the section. Then, by definition

$$\phi = \frac{N}{T} \quad \text{.......................................(11).}$$

Now, T may be approximated by the product of the total tissue area in one of the planes of section and the distance (10 micra) between sections; but N is usually a difficult number to ascertain, due to the difficulty of locating islet centroids with any satisfactory degree of approximation in many instances. As indicated above, however, there are two resources available.

In the first place we may count the total number, H, of islets having any part in the section. This is simple if no islet has more than one discrete part in the section. Otherwise, we may estimate a correction to the total count of islet particles discrete in the section if the necessary data are available for a random sample of these islets.

Now, let η be the number of sections in which any part of a given islet appears. Then we assume that the probability that the centroid of this islet lies in an arbitrary one of these sections is $1/\eta$. Accordingly, if the H islets be numbered from 1 to H and η_j bear the same relation to islet number j in this set as does η to the islet mentioned above, then we may take N', defined by

$$N' = \sum_{j=1}^{j=H} \frac{1}{\eta_j} \quad \text{.....................................(12)},$$

as an estimate of N. Of course, N' may in turn be estimated by means of a random sample of the H islets. Whether such successive approximations are admissible or not must be decided in each case separately.

We have, however, an alternative method of estimating N, if it is admissible to assume that the difference, if any, between the probability that the centroid lies in the arbitrary section and that for any other point (uniquely determined) which we may substitute for the centroid is tolerable; i.e. we agree to tolerate resultant errors of the substitution. In this case the number, H', of the above-mentioned H islets lying at least in part in the section of reference, but not in a given adjoining section, may be taken as an estimate of N. Obviously, as an adjoining section may be chosen in two different ways we may obtain H'' as the homologue of H' for the alternate choice. Accordingly, either H', H'' or $\frac{H'+H''}{2}$ may be taken as an estimate of N.

Now, in a given domain let N be the number of islets belonging to the domain, $\bar{v}$ be their mean volume and I be the sum of their volumes. Let I equal the total volume of islet particles in the domain. This last requirement is special but it may suffice for estimates that it be merely approximated. Approximately in this case and exactly in the ideal case given above, then we have the relations

$$I = N \, . \, \bar{v} \quad \text{.....................................(13)},$$

where I is the total volume of insular tissue in the domain; and if T is the (total tissue) volume of the domain, and $\phi = N/T$ as before, and we let $\delta = I/T$, then, obviously,

$$\delta = \phi \, . \, \bar{v} \quad \text{.....................................(14)}.$$

Accordingly, the estimation of any two of the variables suffices to give an estimation of the third in either (13) or (14), and if we have an estimation of T then all these variables may be estimated. If all three variables in either equation may be estimated independently these relations furnish a means of checking the results.

Summary.

Some of the general aspects of the theory and practice of the quantitative study of geometric properties of microscopic configurations by means of projection have been discussed. Further elaboration is described in the succeeding communication together with applications.

THE GEOMETRIC PROPERTIES OF MICROSCOPIC CONFIGURATIONS. II. INCIDENCE AND VOLUME OF ISLANDS OF LANGERHANS IN THE PANCREAS OF A MONKEY.

By WILLIAM R. THOMPSON AND RAYMOND HUSSEY, WITH THE ASSISTANCE OF JOSEPH T. MATTEIS, WILLIAM C. MEREDITH, GEORGE C. WILSON AND F. ERWIN TRACY.

MANY attempts have been made to study geometric properties of islands of Langerhans in the pancreas; but, as is readily apparent in a critical survey of published results, the field is fraught with pitfalls to trap the unwary. One of these is the tacit assumption that, by proposed methods, random samples of islets in a region of the organ are obtained; another is the failure to take proper account of the difference between counts of whole islets and counts of parts of islets which lie in a given section; and a third is the laying down of criteria for estimation of islet volume which may lead to estimates for the same islet differing by many hundred per cent. dependent upon an arbitrary choice of direction for cross-section, or to similar differences in estimate of volume in the case of two islets of identical size and shape but different orientation with respect to direction of axis of cross-section, or to instances in which the dimensions whose measurement is required may not even exist.

The first, apparently, to avoid these pitfalls was Bensley*. By the nature of his methods, however, he was obliged to confine his attention to making counts of islets in tissue and estimations of the mass of the tissue in which they were contained. In this manner he was able to estimate the number of islets per unit mass in a pancreas as well as the total number of islets in the organ. As pointed out by Bensley*, however, the method he employed is subject to difficulty in application due to the fact that the staining technique is delicate (though yielding excellent contrast when correctly performed), and the time in which counts must be made is limited to a few hours. Clark†, in applying the methods of Bensley to a study of the human pancreas, encountered such difficulties.

As stated in the preceding communication‡, studies of the geometric properties of microscopic configurations, particularly volumetric relations of islands of Langerhans in the pancreas, have been conducted in this laboratory for several years. The previous paper was designed to serve as an introduction to subsequent reports

* Bensley, R. R., *Am. J. Anat.* 1911—12, Vol. XII. pp. 297—388.

† Clark, E., *Anat. Anz.* 1913, Vol. XLIII. pp. 81—94.

‡ Thompson, W. R.: see pp. 21—26 of this issue.

and dealt with general aspects of the subject for which the name *projectometry* was proposed. The present purpose is to present an application of methods suggested therein to a study of the islands of Langerhans in the pancreas of a monkey, and to contrast the results obtained by two independent methods of estimation of mean islet volume in a given region. The data to be employed were obtained from observations made by certain students* working in this laboratory.

Technical Procedures.

Material for such observations was prepared in the following manner:

The pancreas of a presumably normal monkey (*macacus rhesus*) under general anaesthesia was removed and placed without delay in a 10 °/₀ solution of formalin. Under formalin it was cut into twenty blocks and two small portions at either extremity, which latter were excluded in the present observations (the cuts being made approximately at right angles to the major axis of the organ). These blocks were numbered in the order in which they lay originally in the organ from one to twenty, number one block being that nearest to the extremity called the tail. From each of the odd-numbered blocks 120 serial sections were cut, ten micra in thickness, with a precision microtome, numbered in order of section from one to 120 and mounted, the sections so obtained being from a region of the block as near as practicable to the face of the original section of the organ which was nearest to the tail end of the pancreas. The odd numbered sections were stained with haematoxylin and eosin and covered with cover-slips in the usual manner. Accordingly, there were 60 sections so prepared for each of the odd-numbered blocks at approximately equal intervals along the major axis of the organ.

The process of selecting material for observation was as follows:

One section in each set of 60 (corresponding to an original block) was designated as a *master* or principal section of reference. For this purpose section #59 was taken unless it was found to be defective as to preparation to such an extent that proposed operations could not be effected, in which case a suitable section as near as possible to #59 was taken as master. In the experience to be reported #59 was satisfactory in eight of the ten instances. A chart was made of each master section, giving the location of each discrete region of insular tissue in it. A discrete part of an islet lying in a given section will be called an *insular particle*, for convenience in the following discussion; if the section is a *master* then the particle will be called a *master insular particle*; and insular particles belonging to the same islet will be called *associate particles.* Accordingly, in conjunction with the chart for each *master section,* the insular particles belonging to it were enumerated. It should be noted that in such enumeration no significance was attached to whether two given master insular particles were also associate particles or not. This was deliberate and will be taken effectively into account later. The procedure was then to select by lot ten non-associate particles from each master section and to estimate the

* Matteis, J. T., Meredith, W. C., Wilson, G. C., and Tracy, F. E.

volume of each associated islet by the projectometric method previously suggested*, and to use the master section for total tissue and total insular tissue area measurements. Furthermore, in subsequent developments we designate by Z the aggregate number of insular particles in a given master section.

In each instance tissue areas were estimated by means of the projection and tracing of a magnified image with a projectiscope, the measurement of the image area by means of a planimeter, and the estimation of the corresponding mean area magnification. For the case of perfect magnification (i.e. with no distortion) the area magnification is equal to the square of the linear magnification. The projectiscope used in this work, however, is subject to distortion of images (inherent in its design). This distortion, however, is symmetrical with respect to the centre of the circular image field.

The result of a short sequence of observations of mean linear magnification of a straight line segment of given length, situated so that one end, P_0, of its image, $\overline{P_0P_i}$, is at the centre of the image field, is given in Table I and represented

TABLE I.

r in cms.	ρ in cms.
0·001	0·587
0·002	1·175
0·003	1·752
0·004	2·320
0·005	2·915
0·006	3·502
0·007	4·085
0·008	4·672
0·009	5·272
0·010	5·870
0·011	6·482
0·012	7·095
0·013	7·712
0·014	8·342
0·015	8·980
0·016	9·630
0·017	10·272

The values of ρ listed are the means of four observations, r is treated as if exact.

graphically in Text-fig. 1 with a parabola fitted to the appropriately weighted observations of mean linear magnification by the method of least squares. The weighting is proportional to the square of the segment length (which is approximately proportional to the precision of the measurement). We let r be the length of a given segment (as described above) and ρ be the length of its image, $\overline{P_0P_i}$ (both in cms.). Then the linear magnification factor, f, is given by

$$f = \frac{\rho}{r} \quad \dots\dots\dots\dots\dots\dots\dots\dots (1);$$

* Thompson, W. R., *loc. cit.*

and the above-mentioned parabola is given by

$$f = kr^2 + h \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2),$$

where k and h are constants determined by the method of curve fitting indicated above. Now it may be verified readily that the curve defined by

$$\rho = kr^3 + hr \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3),$$

if fitted to the original data (as given in Table I) by the simple (equal weight) method of least squares, gives exactly the same values to k and h as in (2); and (1), (2) and (3) then form a consistent set of relations. By such calculation the approximations, $h \backsim 579{\cdot}95$ and $k \backsim 8{\cdot}201\,(10)^4\,.\,\text{cm.}^{-2}$, were obtained.

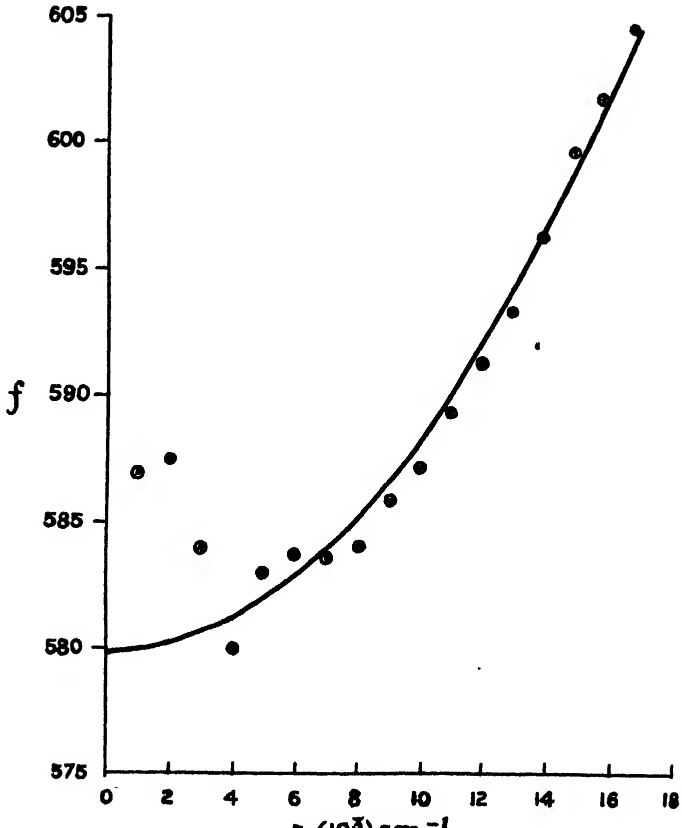


Fig. 1.

However, the regions whose areas are to be estimated are irregular; and, accordingly, a strict application of the information thus obtained is quite difficult. In the case of islet measurements it is possible to place image areas so that their centroid is approximately in the centre of the image field and roughly approximate their areas by similarly placed circles whose area magnification would be f^2 if the point P_i lie on the circumference of the circle in the image. In the work to be

reported at present f was replaced by an approximation, F, taken for $r=0{\cdot}01$ cm., and the area magnification in such instances was approximated by F^2. A rough estimate of the error thus involved may be made as follows, by estimating the maximum error for the approximating circles' magnification so introduced.

From the data (given in the graph) we have that in the field

$$579{\cdot}9 \leqslant f \leqslant 603{\cdot}7 \quad \text{and} \quad F=588{\cdot}15\ldots \qquad (4).$$

Now, let $E=|F-f|$, then

$$E \leqslant 15{\cdot}6 \qquad (5),$$

and the maximum relative error,

$$\frac{E}{F} \leqslant \frac{15{\cdot}6}{588{\cdot}1} < 0{\cdot}027 = 2{\cdot}7\ {}^\circ/_\circ.$$

Accordingly, the relative error in estimating the magnification of these circular areas cannot exceed $5{\cdot}5\ {}^\circ/_\circ$, and this is taken as an indication of the greatest errors to be expected in the present applications due to replacement of the variable f by the constant F in our calculations. A similar approximation was used in the case of total tissue area measurements.

From the data obtained in the measurement of insular areas of a given islet in the successive odd-numbered sections a graph may be made by location of points in a Cartesian co-ordinate system having x, the distance in micra (μ) from an arbitrary point to the plane of section in a given instance, as abscissa, and y, the total area in square micra (μ^2) of the cross-section of the given islet in that plane, as ordinate; and drawing a smooth curve (or a polygon) to represent these points in the conventional manner.

Now, let u be the volume of tissue in cubic micra (μ^3) to which a unit square in this co-ordinate system corresponds; K be the area of such a unit square and (in the same units) G be the area (of the connex) bounded by the curve mentioned above and the axis of abscissae. Now, let v be defined by

$$v=u\,.\frac{G}{K} \qquad (6).$$

Then v is the estimate of the islet volume in μ^3 which is sought. Obviously, instead of the co-ordinates used above, any pair (Ax, By), where A and B are known constants, may be used, whereupon the variables defined above are related by

$$v=\frac{uG}{ABK} \qquad (7).$$

The areas mentioned above are estimated by means of a planimeter in the usual manner. If a planimeter graduated in square inches be used it may be convenient to adopt observed image areas as ordinates (By), then, if F^2 be the area magnification factor, we have

$$B=\frac{F^2}{(2{\cdot}54)^2}\,.\,10^{-8}.$$

Obviously, the calibration of instruments is not to be overlooked; and correction factors should be applied to readings wherever their neglect would lead to errors not to be tolerated. In the work herein reported most of these elements have been examined, but attention has been focused primarily upon investigation of reproducibility and consistence rather than upon absolute reliability, although the only major element of the various technical operations involved not to be so examined is that of the reliability of the microtome. The instrument used, however, was of the type known as the *precision microtome.* The precision of the method of volume estimation by successive mechanical integration as described above is indicated by the following data. Sections 5 micra in thickness were employed instead of 10 micra thick sections as specified above, but these were arranged into four sets of serial sections at intervals of 20 micra. Thus it was possible to obtain four otherwise independent estimates of the volume of the same islet by the method in question with this slight alteration. The results were as follows:

Set No.	$v\,(10)^{-5}.\,\mu^{-3}$	Deviation
1	19·06	1·62
2	17·35	− ·09
3	17·48	·04
4	15·87	− 1·57

Mean = 17·44; Mean absolute Deviation = ·83.

These four observations, however, were made upon an islet of a human pancreas; but, in view of similarities of structure, the result is taken as indicative of what may be expected in the case of an islet of a monkey's pancreas. The a.d. is the mean absolute deviation from the mean of the observed values of v, of which it is approximately 4·8 °/$_{\circ}$.

For a given islet (in accord with the definition given above *) let α be the number of associate particles in the corresponding master section, and let η be the number of sections in the original set as cut (120 serial sections, 10 micra in thickness) in which any part of the islet appears. For 100 islets in the monkey's pancreas (ten from each of ten blocks, as described) the estimates of v obtained by the method described above are given in Table II together with the corresponding values of an approximation of $\frac{\eta}{2}$ by the number of sections at intervals of 20 micra (the stained sections) in which any part of the given islet appears. In each instance in the present experience $\alpha = 1$.

We employ $M_{(v)}$, a weighted mean of observed values of v, as an estimate of mean islet volume in a given region; the weighting being proportional to the reciprocal of the *à priori* probability of selection, in accord with the previous communication *.

* Thompson, W. R., *loc. cit.*

In order to facilitate such estimation let us assume that we may proceed (with only negligible differences in approximations) as if the following ideal situation existed:

The organ was originally sectioned by a set of parallel planes at intervals of 10 micra, of which the sections actually obtained in the above-mentioned experience would be a part; and that the *à priori* probability p of selection of a given islet is the same as if the master section were chosen at random from the complete set, an insular particle of which was then chosen at random and the islet of which it is a part considered so to be selected.

TABLE II.

v in $10^4 . \mu^3$	$\frac{\eta}{2}$	v in $10^4 . \mu^3$	$\frac{\eta}{2}$	v in $10^4 . \mu^3$	$\frac{\eta}{2}$	v in $10^4 . \mu^3$	$\frac{\eta}{2}$	v in $10^4 . \mu^3$	$\frac{\eta}{2}$
(Block #1)		(Block #5)		(Block #9)		(Block #13)		(Block #17)	
382·1	13	28·2	4	135·0	7	110·7	10	227·1	8
99·3	8	392·9	9	24·3	4	18·6	5	17·8	4
24·5	4	192·8	11	22·3	4	8·1	3	49·1	6
12·8	3	19·7	4	26·7	3	4·1	2	17·2	4
47·1	5	7·9	3	49·5	6	44·3	6	26·6	6
204·1	7	32·0	5	53·2	6	13·4	5	12·3	3
32·9	4	228·7	13	232·1	9	8·7	3	9·8	2
14·2	3	172·0	8	150·8	8	10·5	3	93·5	7
45·3	7	40·9	6	65·2	7	40·7	6	31·2	5
65·5	4	223·9	7	168·1	7	44·0	5	412·4	10
(Block #3)		(Block #7)		(Block #11)		(Block #15)		(Block #19)	
202·4	8	90·3	5	39·7	4	39·3	5	23·3	3
10·5	4	51·9	5	6·4	2	6·8	3	129·2	8
17·9	4	69·6	6	7·3	3	42·8	5	162·8	8
7·8	3	39·0	3	24·6	6	8·7	2	107·7	9
25·3	5	58·6	7	39·5	4	10·3	2	14·2	3
27·2	4	38·0	8	140·3	7	6·4	3	32·7	5
25·5	4	48·3	4	43·7	5	25·3	5	63·9	8
148·6	8	116·0	8	84·6	6	10·3	3	796·5	13
153·8	9	24·2	3	571·2	12	11·8	3	7·8	4
382·6	13	40·5	3	205·9	8	10·4	2	83·0	7

Now, let η be the number of sections of the set in which a part of the given islet appears, α be the number of associate particles of this islet in a given section, and Z be the number of insular particles in this section. Then it can be demonstrated rigorously that there exists a number λ (the same for all islets in the set) such that

$$\lambda p = \Sigma \frac{\alpha}{Z} \quad \text{.......................................(8)},$$

where the summation is over the complete set of sections, except where $Z=0$, if ever.

However, the terms in this expression will all be zero except exactly η of them. Accordingly, by the expression for the weighted mean as given in (3) of the previous paper* we have

$$M_{(v)} = \frac{\Sigma f . \frac{v}{p}}{\Sigma \frac{f}{p}} \quad \text{.................................(9)}$$

(where the summation is over the entire number of islets in the organ and f is the number of times a given islet has been selected as described previously*), p need not be known wherever $f = 0$ and λp may be substituted for p, obviously with the identical result. Furthermore, in the actual experience $f = 1$ or 0 in every instance†.

Now, if in (8) it is tolerable (as is assumed in the present experience) to replace $\Sigma \frac{a}{Z}$ by the approximation $\frac{a}{Z} . \eta$, where we take the values of a and Z for some particular one of the η sections where $a \neq 0$ such as the *master*; then we have for $M_{(v)}$ the expression

$$M_{(v)} = \frac{\Sigma \frac{Z . v}{a . \eta} . f}{\Sigma \frac{Z}{a . \eta} . f} \quad \text{.............................(10),}$$

where the summation is over the islets actually selected. In the present experience, as has been stated, $f = a = 1$, and *for the means for a given block*, as Z is the same for the ten islets measured, it may therefore be eliminated from the formula. However, this is not permissible in the estimate of the weighted mean volume for the whole organ, for which, by (10) and Table II, we have (from the values of Z given in Table III)

$$M_{(v)} = 5{\cdot}66\,(10)^5 . \mu^3 \quad \text{.......................... (11).}$$

TABLE III.

Master Section of Block No.	I (in $10^8 . \mu^3$)	T (in $10^9 . \mu^3$)	Z	ω	N (approx. by $Z . \omega$)	δ $\left(\frac{I}{T}\right)$	ϕ (in islets per mm.^{3})
1	8·60	·287	157	·1052	16·5	·0300	57·5
3	8·18	·272	134	·0986	13·2	·0301	48·5
5	8·85	·246	128	·0874	11·2	·0360	45·5
7	5·01	·250	86	·1105	9·5	·0201	38·0
9	7·97	·264	148	·0916	13·6	·0302	51·5
11	7·13	·335	150	·1109	16·6	·0213	49·6
13	6·23	·295	119	·1267	15·1	·0211	51·2
15	7·28	·343	128	·1716	22·0	·0212	64·2
17	5·94	·313	116	·1118	13·0	·0190	41·5
19	10·30	·466	289	·0911	26·3	·0221	56·5
Totals	75·49	3·071			157·0	$\delta_0 = {\cdot}0246$,	$\phi_0 = 51{\cdot}1$

* Thompson, W. R., *loc. cit.*

† The neglect of the probability of multiple selection of the same islet in this manner is obviously tolerable here.

For the sake of brevity in the following discourse we let $V = M_{(v)}$ as estimated for a given block, the values obtained being listed in Table IV.

TABLE IV.

Mean islet volume estimates for ten regions of a pancreas of a monkey, contrasting two technical methods of estimation as described in the text.

Block No.	V (in $10^5 . \mu^3$)	V' (in $10^5 . \mu^3$)	$\frac{V - V'}{V + V'}$	$\frac{V}{V'}$
1	6·01	5·21	·07	1·15
3	6·00	6·20	−·02	0·97
5	9·16	7·90	·07	1·16
7	5·18	5·27	−·01	0·98
9	7·37	5·86	·11	1·26
11	6·57	4·29	·21	1·53
13	2·04	4·13	−·34	0·49
15	1·40	3·31	−·41	0·42
17	5·35	4·57	·08	1·17
19	8·28	3·92	·36	2·11
Arithmetic Means	5·74	5·07	·01	(Geometric Mean = 1·02)

Now, let us turn our attention to certain other functions of this material. Let I be the total volume of insular tissue in a given region, and let T be the total tissue volume in the same region; then we define δ by

$$\delta = \frac{I}{T} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12),$$

the ratio of insular to total tissue volume in the region. The observed values of I, T and δ for the master sections of the ten blocks are given in Table III, where tissue volumes are estimated by the product of the thickness of the section (10μ) and the corresponding estimate of tissue area (projectometrically) in the section. In the same table are given the corresponding values of ω, defined as the mean reciprocal of η for the ten islets of the block previously selected for volume measurement.

As indicated in the previous paper*, we may approximate N, the number of islets belonging to the master section† (in the sense there defined), by $Z . \omega$; and ϕ, the population density, by $\frac{Z . \omega}{T}$. The values of these variables are given in Table III also.

* Thompson, W. R., *loc. cit.*

† Neglecting the difference between Z and H (as $a = 1$ in each of 100 instances).

By means of these data we may obtain an independent estimate of mean islet volume in these regions, V', given by

$$V' = \frac{\delta}{\phi} \sim \frac{I}{Z.\omega} \quad \text{.............................(13)};$$

according to Equations (13) and (14) of the previous communication *. In Table IV are given the corresponding values of V' together with those of V and the functions of these variables defined by

$$D = D_{(V,V')} = \frac{V - V'}{V + V'}, \quad \text{and} \quad R = R_{(V,V')} = \frac{V}{V'} \quad \text{...........(14)}.$$

The mean of the values of D in the ten instances is 0·01 with mean a.d. = 0·17; and the geometric mean of R is 1·02, which means approximate the ideal values (zero in the first and unity in the second instance) much better than it seems reasonable to expect in view of known inherent errors of the method. Indeed, it seems improbable that another experience of this kind would lead to such close approximation, deviations of 0·05 or more from the ideal appearing more likely than not, although more precise results should be obtainable from more extensive investigations, wherein some of the roughness of successive approximations might be eliminated. As has been stated, however, the object of the present paper is to indicate methods by means of which such results may be obtained, employing data at present available merely for illustrative purposes.

It may be noted that the function, $D_{(V,V')}$, is merely the deviation of V from the mean of V and V' divided by this mean, i.e.

$$\frac{V - \frac{V + V'}{2}}{\frac{V + V'}{2}} = \frac{V - V'}{V + V'} = D_{(V,V')} \quad \text{.......................(15)};$$

and, obviously, $\qquad D_{(V',V)} = - D_{(V,V')}$.

Now, in a manner similar to that in which V' is taken as an estimate of mean islet volume in the block regions we may obtain V_0' as an estimate of mean islet volume in the organ which may be compared with the estimate $M_{(v)}$ given in (11).

We obtain an estimate δ_0 of the ratio of insular to total tissue volume in the organ by

$$\delta_0 = \frac{\Sigma I}{\Sigma T} \quad \text{..................................(16)},$$

where the summation is over the ten values in each instance as given in Table III. Similarly we obtain ϕ_0, an estimate of population density for the whole organ, given by

$$\phi_0 = \frac{\Sigma Z.\omega}{\Sigma T} \quad \text{..................................(17)}.$$

* Thompson, W. R., *loc. cit.*

Then, in accord with previous procedure, we define V_0', the estimate of mean islet volume in the organ, by

$$V_0' = \frac{\delta_0}{\phi_0} = \frac{\Sigma I}{\Sigma Z \,.\, \omega} \quad \text{.............................(18).}$$

From the data the value, $V_0' = 4{\cdot}81\,(10)^5 \,.\, \mu^3$, was obtained, which approximates $M_{(v)}$ as given in (11) as closely as may reasonably be expected in view of the roughness of the successive approximations (e.g. the approximation of η by an *even* integer) and the great dispersion in islet volume which is obvious. Both are approximations of $\tilde{v}$, the mean of the volumes of all islets in the organ.

Discussion.

The roughness of the estimation of η by an even integer need not be a feature of subsequent work; and, indeed, the approximation of λp in (8) by $\frac{\alpha}{Z} . \eta$ is not necessary as the exact value of $\lambda p = \Sigma \frac{\alpha}{Z}$ may be calculated in each instance if the several adjacent sections in which the measured islets have a part are examined for the data required under (8). Furthermore, we may use the method of estimation of N by count (as suggested in the previous paper*) of islets having a part in the master section but not in a given adjoining section or preferably by the mean of the two estimates possible thus by alternative choice of an adjoining section. In this manner we may obtain estimates of δ, ϕ and mean islet volume $\tilde{v}$ in a region from only three successive sections instead of 120 as here employed, only 60 of which were stained however, thus reducing the cost and labour of preparation to less than 5 °/ₒ of that in the present work in this regard.

However, studies of probable dispersion including class-frequency relations would not so be possible, but the treatment of this phase of the subject will be left for another communication.

Were the distances along an axis and the direction of sectioning planes at given points known (as could easily have been the case were the need appreciated at the time of cutting the blocks and sections) then it would be a simple matter to estimate total organ volume and total insular tissue volume by successive mechanical integration as in the case of islet volume; and, indeed, to estimate the total number of islets in the organ.

The danger of *ad hoc* calculations has been kept well in mind throughout the present work. The theory upon which calculations are based contains no arbitrary constants fixed by design to attain the end result, and was elaborated without regard to the result in this particular instance. A somewhat more crude theory, first evolved, led to almost exactly the same result as given here, the agreement between the corresponding alternate methods of approximation of mean islet volume for the whole organ being even closer. Subsequently introduced refinements improved the theory but altered the resulting estimates of mean islet volume only to a negligible extent in the present experience, which was as expected.

* Thompson, W. R., *loc. cit.*

Summary.

Methods have been presented and their application illustrated from material from the pancreas of a monkey (*macacus rhesus*) whereby $\tilde{v}$, the mean volume of islands of Langerhans in a given region, may be estimated as well as ϕ, the mean number of islets per unit volume of containing tissue, and δ, the ratio of insular to total tissue volume. An alternative method of estimation of $\tilde{v}$ based upon islet volume measurement and selection probability estimates in the case of non-random samples is treated in like manner, giving agreement as good as is to be expected from the extent and precision of available data.

Thus, the two respective estimates of mean islet volume in the pancreas of the monkey to which reference has been made were $4{\cdot}81\,(10)^5 . \mu^3$ and $5{\cdot}66\,(10)^5 . \mu^3$; and the mean value of the ratio of insular to total tissue was found to be 0·0246, and the mean number of islets per cubic millimeter was approximately 51.

Methods of estimation of total insular volume, total tissue volume and total number of islands in the organ are suggested although the necessary data are not available (in the present experience) for such calculations.

The influence of distortion in projection has been studied and methods of restricting errors so introduced in microscopic area measurement have been presented.

A BESSEL FUNCTION DISTRIBUTION.

By A. T. McKAY, M.Sc.

§ 1. *Introduction.*

As the study of the distributions of the parameters of simple parent populations proceeds, the existence of new and often very complicated distributions is brought to light. These distributions, however, are sometimes only partially known in the sense that several of their positive moments or semi-invariants have, after much laborious effort, been determined. Naturally, the lack of explicit knowledge as to the precise functional expression of the distribution is not by any means entirely remedied by a mere fitting of the Pearson Type Curves, for there can possibly exist serious discrepancies between the fifth and higher moments of the fitting and the fitted curve. In view of this it would seem most desirable that the family of known distribution curves, i.e. curves whose functional form and all semi-invariants are known, which at present consist chiefly of the Pearson Curves, should be added to wherever possible. It is the purpose of the present paper, therefore, to bring to notice a most interesting Bessel Function distribution, the semi-invariants of which are all given in a surprisingly simple form.

Further, it is known that the Pearson Type Curves do not always provide an entirely satisfactory fit to data, and it is therefore very useful to have at hand alternative fitting curves which can be tried on such experience. It is hoped that the Bessel Function distribution given herein may prove of service in fitting data which have Betas below the Pearson Type III line but which do not prove amenable to a fitting by a Pearson Curve.

§ 2. *The Distribution.*

The distribution which we are to consider is

$$y = y_0 e^{-cx/b} \, . \, |x|^m \left\{ \begin{array}{c} \pi I_m \, |x/b| \\ \text{or} \\ K_m \, |x/b| \end{array} \right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1),$$

where I_m and K_m are the Bessel Functions of the second kind as defined by Watson*. The upper function must be employed when $|c| > 1$, in which case the distribution curve extends from 0 to ∞ if c is positive or from 0 to $-\infty$ if c is negative. When $|c| < 1$ the lower function is to be employed, in which case the distribution curve extends from $-\infty$ to $+\infty$. The quantity b is a positive constant and $(m + \frac{1}{2}) > 0$. The value of y_0 is given by

$$y_0 = \frac{|1 - c^2|^{m+\frac{1}{2}}}{\pi^{\frac{1}{2}} 2^m b^{m+1} \Gamma(m + \frac{1}{2})} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2).$$

* G. N. Watson, *Theory of Bessel Functions.*

For the purpose of determining the moments of the distribution it is most convenient to discuss the two cases separately, viz.

$$\text{Case (i): } |c| > 1,$$
$$\text{Case (ii): } |c| < 1.$$

It should be noted that in equation (1) only the index of the exponential term changes sign with x.

§ 3. *Determination of Moments.*

Case (i)*.

Consider the expression

$$E = \int_0^\infty e^{(\alpha - c/b)x} . x^m . I_m(x/b)\, dx \quad \text{.......................(3)},$$

where α is a quantity entirely at our disposal and c/b is positive. Now

$$I_m(x/b) = \sum_0^\infty (x/2b)^{m+2r} / \Gamma(r+1) . \Gamma(m+r+1) \quad \text{..............(4)};$$

hence substituting in (3) we have after integration

$$E = \sum_0^\infty \Gamma(2m+2r+1) / \Gamma(r+1) . \Gamma(m+r+1) . (2b)^{m+2r} . (c/b - \alpha)^{2m+2r+1} \quad \text{.........(5)}.$$

Now the Gamma-Function duplication formula is

$$2^{2z-1}\, \Gamma(z) . \Gamma(z + \tfrac{1}{2}) = \pi^{\frac{1}{2}}\, \Gamma(2z) \quad \text{.......................(6)},$$

and writing $z = m + r + \frac{1}{2}$ in the latter we find

$$2^{2m+2r}\, \Gamma(m + r + \tfrac{1}{2}) . \Gamma(m + r + 1) = \pi^{\frac{1}{2}}\, \Gamma(2m + 2r + 1) \quad \text{.........(7)},$$

whence equation (5) reduces to

$$E = \frac{(2b)^{m+1}}{2\pi^{\frac{1}{2}}} . \sum_0^\infty \Gamma(m + r + \tfrac{1}{2}) / \Gamma(r+1) . (c - \alpha b)^{2m+2r+1} \quad \text{..............(8)}$$

$$= (2b)^{m+1} . \Gamma(m + \tfrac{1}{2}) / 2\pi^{\frac{1}{2}} ((c - \alpha b)^2 - 1)^{m+\frac{1}{2}} \text{ since } |c| > 1 \quad \text{.........(9)}.$$

We conclude therefore that if μ_r is the rth positive moment of the distribution about the origin, then μ_r is the coefficient of $\alpha^r r!$ in the expression $\pi y_0 E$, that is in the expression

$$E_1 = [(1 - c^2) / (1 - (c - \alpha b)^2]^{m+\frac{1}{2}} \quad \text{.......................(10)}.$$

Case (ii).

Consider the expression

$$E' = \int_{-\infty}^\infty e^{hx} |x|^m K_m |x/b|\, dx = b^{m+1} \int_0^\infty (e^{bhx} + e^{-bhx}) x^m K_m(x)\, dx, \text{ where } |bh| < 1 \quad \text{.........(11)}.$$

Now it can be shown† that

$$\int_0^\infty e^{bhx} x^m K_m(x)\, dx = (\tfrac{1}{2}\pi)^{\frac{1}{2}}\, \Gamma(2m+1) . \frac{F(\frac{1}{2} - m, m + \frac{1}{2}, m + \frac{3}{2}; (1 + bh)/2)}{\Gamma(m + \frac{3}{2})\ (1 - hb)^{m+\frac{1}{2}}} \quad \text{.........(12)},$$

* The device used here of multiplying by $e^{\alpha x}$ and integrating or summing often proves most useful in dealing with distributions which have all the positive moments finite. For instance the method leads to the following neat form for the semi-invariants for the Binomial Distribution $(q+p)^n$. The rth semi-invariant $= n . \log(1 + p\Delta) . 0^r$. Since the differences have been tabulated, or in any case are quite easily found, the formula is a specially useful form.

† G. N. Watson, *loc. cit.*

where F is the usual hypergeometric function notation. Whence we have

$$E' = (\tfrac{1}{2}\pi)^{\frac{1}{2}} b^{m+1} 2^{m+\frac{1}{2}} \Gamma(2m+1) . \frac{[z^{m+\frac{1}{2}} F_1(z) + (1-z)^{m+\frac{1}{2}} F_1(1-z)]}{\Gamma(m+\frac{3}{2}) . (1-b^2h^2)^{m+\frac{1}{2}}} \quad \text{...(13)},$$

where $z = (1+bh)/2$ and $F_1(z)$ is the F-term of equation (12). Let us call the quantity in square brackets G and consider it thus:

It has been shown† that provided $0 < z < 1$ that

$$\begin{aligned} &\Gamma(c-a) . \Gamma(c-b) . \Gamma(a) . \Gamma(b) . F(a, b, c; z) \\ &\quad = \Gamma(c) . \Gamma(a) . \Gamma(b) . \Gamma(c-a-b) . F(a, b, a+b-c+1; 1-z) \\ &\quad\quad + \Gamma(c) . \Gamma(c-a) . \Gamma(c-b) . \Gamma(a+b-c) . (1-z)^{c-a-b} \\ &\quad\quad\quad \times F(c-a, c-b, c-a-b+1; 1-z) \quad \text{..................(14)}, \end{aligned}$$

whence substituting in this formula the appropriate values we find

$$F_1(z) = A . z^{-(m+\frac{1}{2})} - (1-z)^{m+\frac{1}{2}} F_2(1-z) \quad \text{..................(15)},$$

where $$A = \Gamma(m+\tfrac{3}{2}) . \Gamma(m+\tfrac{1}{2}) / \Gamma(2m+1)$$

and $$F_2(z) = F(2m+1, 1, m+\tfrac{3}{2}; z).$$

Now applying (14) to expand $F_2(z)$ we derive

$$F_2(z) = -F_2(1-z) + A/[z(1-z)]^{m+\frac{1}{2}} \quad \text{..................(16)}.$$

So from (13), (15) and (16) we find that $G = A$ and is independent of z. Whence Equation (11) reduces to

$$E' = (2b)^{m+1} \pi^{\frac{1}{2}} \Gamma(m+\tfrac{1}{2}) / 2(1-b^2h^2)^{m+\frac{1}{2}} \quad \text{..................(17)}.$$

Writing $h = (a - c/b)$ and multiplying by y_0 we finally deduce as in Case (i) that μ_r is the coefficient of $\alpha^r r!$ in the expression E_1 given in Equation (10).

§ 4. *Semi-invariants and Criteria.*

We conclude then from Equation (10) that the rth moment about the origin of the distribution is

$$\mu_r = \frac{d^r}{d\alpha^r} \left[\frac{1-c^2}{1-(c-b\alpha)^2} \right]^{m+\frac{1}{2}}_{\alpha=0} \quad \text{..........................(18)}.$$

Whence if k_r is the rth semi-invariant we deduce

$$k_r = (r-1)! . (m+\tfrac{1}{2}) . \frac{b^r[(c-1)^r + (c+1)^r]}{(c^2-1)^r} \quad \text{..............(19)}.$$

Hence we have $$\text{Mean} = \bar{x} = (2m+1)\, bc/(c^2-1) \quad \text{..............................(20)},$$

$$\sigma_x^2 = (2m+1)\, b^2 (c^2+1)/(c^2-1)^2 \quad \text{....................(21)},$$

$$\beta_1 = 4c^2(c^2+3)^2/(2m+1)(c^2+1)^3 \quad \text{....................(22)},$$

$$\beta_2 = 3 + 6(c^4+6c^2+1)/(2m+1)(c^2+1)^2 \quad \text{...........(23)},$$

$$\text{Criterion } \kappa_1 = 12(c^2-1)^2/(2m+1)(c^2+1)^3 \quad \text{....................(24)}.$$

From Equation (21) we see the necessity for the restriction that $(m+\frac{1}{2}) > 0$, for the standard deviation must be positive.

† E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*.

§ 5. *Properties of the Distribution.*

Writing $\delta = (\beta_2 - 3)/\beta_1$ and $c^2 = t$, we find from Equations (22) and (23)

$$2t(t+3)^2\delta - 3(t+1)(t^2+6t+1) = 0 \quad\text{.....................(25),}$$

which, when δ is given, is a cubic equation to find positive values of t. Further, from Equation (24), the criterion κ_1 is essentially positive, hence the distribution is leptokurtic and $\delta \geqslant 1{\cdot}5$. The lines $\beta_2 = 1{\cdot}5\beta_1 + 3$ (the Pearson Type III Line) and $\beta_2 = 1{\cdot}57735\beta_1 + 3$, which can be called the "Bessel Line," are thus critical lines for the distribution. Unfortunately, these lines are too close together to be shown conveniently on a scaled diagram, but the following diagram shows their significance, besides indicating the occurrence of the roots of Equation (25). For a given value of δ the smallest positive root of (25) is t_1, the medium root t_2, and the largest root t_3.

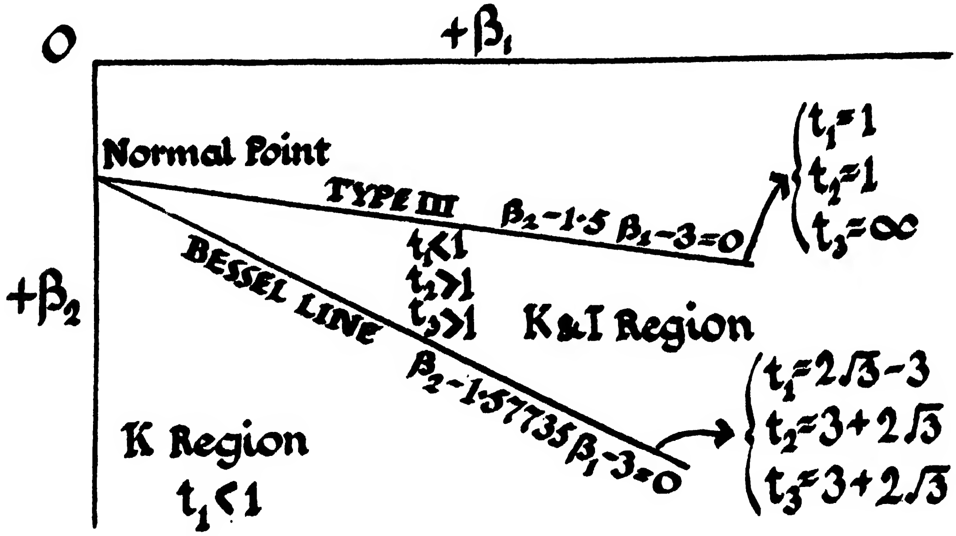


Table I gives the values of t_1, t_2 and t_3 for a series of values of δ corresponding to the overlapping *K-I* region, while Table II gives the value of t_1 (the only root) corresponding to the *K* region.

For values of δ greater than 3·00, the approximation

$$t_1 = x_0 = (6\delta - 7)/\{(6\delta - 7)^2 + (4\delta - 7)\} \quad\text{.....................(26)}$$

can be used. The error in using this formula is less than one in a thousand, though if greater accuracy is required the iteration formula

$$x_{n+1} = \frac{3}{(2\delta - 3)}\left\{\frac{x_n - 1}{x_n + 3}\right\}^2 \quad\text{..........................(27)}$$

readily produces the result to any degree of accuracy.

It may be noted that several well-known distributions can be obtained as special cases by choosing the constants suitably.

The Normal Distribution. Put $c = 0$, then, after writing $(2m+1)b^2 = \sigma^2$, make $b = 0$.

TABLE I.

Positive roots of the Equation (25) for values of δ between 1·500 and 1·57735 at intervals of 0·002.

δ	t_1	t_2	t_3
1·500	1·00000	1·0000	∞
1·502	·86837	1·1641	741·97
1·504	·82119	1·2445	366·93
1·506	·78744	1·3125	241·90
1·508	·76050	1·3746	179·36
1·510	·73784	1·4334	141·83
1·512	·71817	1·4903	116·79
1·514	·70072	1·5461	98·896
1·516	·68501	1·6014	85·464
1·518	·67069	1·6565	75·006
1·520	·65753	1·7118	66·631
1·522	·64535	1·7677	59·769
1·524	·63399	1·8284	54·042
1·526	·62336	1·8816	49·187
1·528	·61336	1·9401	45·018
1·530	·60392	2·0000	41·396
1·532	·59498	2·0614	38·219
1·534	·58648	2·1246	35·407
1·536	·57839	2·1897	32·899
1·538	·57067	2·2571	30·646
1·540	·56328	2·3270	28·610
1·542	·55620	2·3996	26·758
1·544	·54941	2·4755	25·066
1·546	·54287	2·5548	23·511
1·548	·53658	2·6382	22·075
1·550	·53051	2·7261	20·743
1·552	·52465	2·8192	19·502
1·554	·51899	2·9183	18·341
1·556	·51352	3·0242	17·248
1·558	·50821	3·1382	16·216
1·560	·50307	3·2619	15·235
1·562	·49808	3·3971	14·298
1·564	·49324	3·5467	13·398
1·566	·48853	3·7145	12·524
1·568	·48396	3·9062	11·669
1·570	·47950	4·1309	10·818
1·572	·47517	4·4051	9·953
1·574	·47094	4·7634	9·036
1·576	·46682	5·3142	7·956
1·57735	·46410	6·4641	6·4641

Pearson Type III Curve. Write b/c = constant and then make $c=\infty$. Another case arises by writing $b=(c^2-1)\times$ a constant and then putting $c=1$.

The First Product Moment Coefficient in Samples of n drawn from an indefinitely large Normal Population.

Writing $c=-\rho$, $m=n/2-1$, and $x\doteq v$, we derive the distribution as found by Pearson, Jeffery and Elderton*. In this most interesting paper the first four moments and the resulting criteria have been determined. Since the curve for

* Karl Pearson, G. B. Jeffery and Ethel M. Elderton, *Biometrika*, Vol. xxi.

TABLE II.

The positive root of the Equation (25) for values of δ between 1·58 and 3·00 at intervals of 0·02.

δ	t_1	δ	t_1
1·58	0·45889	2·30	0·14046
1·60	·42418	2·32	·13802
1·62	·39577	2·34	·13568
1·64	·37184	2·36	·13341
1·66	·35128	2·38	·13122
1·68	·33333	2·40	·12910
1·70	·31747	2·42	·12706
1·72	·30330	2·44	·12508
1·74	·29054	2·46	·12316
1·76	·27897	2·48	·12130
1·78	·26841	2·50	·11950
1·80	·25872	2·52	·11776
1·82	·24980	2·54	·11606
1·84	·24154	2·56	·11442
1·86	·23386	2·58	·11282
1·88	·22671	2·60	·11127
1·90	·22003	2·62	·10976
1·92	·21376	2·64	·10829
1·94	·20787	2·66	·10686
1·96	·20232	2·68	·10547
1·98	·19709	2·70	·10412
2·00	·19214	2·72	·10280
2·02	·18745	2·74	·10152
2·04	·18300	2·76	·10027
2·06	·17878	2·78	·099045
2·08	·17475	2·80	·097854
2·10	·17091	2·82	·096692
2·12	·16725	2·84	·095559
2·14	·16375	2·86	·094451
2·16	·16040	2·88	·093370
2·18	·15719	2·90	·092314
2·20	·15412	2·92	·091282
2·22	·15117	2·94	·090273
2·24	·14833	2·96	·089287
2·26	·14561	2·98	·088323
2·28	·14299	3·00	·087378

Case (ii), that is $|c| < 1$, is but slightly more general (for m is not more restricted than that $(m+\frac{1}{2}) > 0$) than that considered by the above authors, the reader is recommended to refer to this paper, where he will find a selection of curves and a detailed discussion of some of their characteristics.

The Distribution of the Mean of Samples of n drawn from an indefinitely large Exponential Distribution, the equation of which is given by

$$y = \tfrac{1}{2}k^2 e^{-k^2|x|}.$$

Write $c = 0$, $m+\frac{1}{2} = n$ and $b = 1/nk^2$. It may be noted that in this case the Bessel Functions degenerate into a rather simpler form, since the order is integer minus a half.

ON THE NORMALITY OR WANT OF NORMALITY IN THE FREQUENCY DISTRIBUTIONS OF CRANIAL MEASUREMENTS.

By E. M. ELDERTON, D.Sc. and T. L. WOO, Ph.D.

(1) It is well known that the general appearance of the distributions of anthropometric characters led Quetelet and afterwards Galton to the assumption that such characters were normally distributed. The normal curve was then introduced into anthropometric discussions, and became almost as much a fetish in anthropometry as in the theory of astronomical observations.

It is, however, undoubtedly true that a number of anthropometric characters, if taken in *not too large samples*, roughly obey the normal law. With larger and larger populations the deviation becomes more and more obvious.

Some recent craniometric investigations suggested that the approximate normality of a considerable number of anthropometric characters might be due to such characters depending upon a variety of elements of growth, and when the number of such elements was reduced there might be a distinct weakening in the normality of distribution. Thus while the stature of adults in samples of, say, a thousand may be approximately normal, cubit is less so, and head length or breadth still less so, although these latter cover several bones of the skull. This is more or less in accordance with the idea that the normality of a distribution arises from the action of a multiplicity of a large number of contributory causes each supplying a small amount of the variation.

(2) It seemed possible to test this conception on a number of measurements on the bones of the skull made by Dr T. L. Woo. The measurements taken are those referred to in Dr Woo's paper "On the Asymmetry of the Human Skull."* The distributions of 50 measurements on single bones of the skull were found and the β_1 (or $\sqrt{\beta_1}$) and β_2 of these fifty cases were computed by Dr Woo. The problem therefore arises as to whether these 50 values of $\sqrt{\beta_1}$ and β_2 can be reasonably treated as those of random samples from a system of normally distributed variates. Unfortunately Dr Woo's β's suffer from two defects when used for our present purpose: (i) They are obtained from populations which vary in size from 451 to 887 individuals, the average size being 777, much nearer the upper than the lower limit; and (ii) they are for 25 homologous characters measured right and left. There is a high correlation between the β's for the characters on the homologous bones, which in itself seems to indicate that the β's are not random deviations from the normal values (0, and 3), but depend upon something characteristic of the bone on which the homologous measurements are taken. The correlation of the right and

* *Biometrika*, Vol. XXII. p. 324 *et seq.*

left $\sqrt{\beta_1}$'s is ·8474 and of the right and left β_2's is ·6258. These correlations between left and right skewness and kurtosis seem in themselves to indicate that skewness and kurtosis are individual to the bone, and that the general system of β's as found by Dr Woo cannot be treated as due to random sampling. Thus the second defect in the system of measurements is not without compensating advantage. The first defect appeared originally to be more serious than it really is. The distribution constants of $\sqrt{\beta_1}$ and β_2 on the hypothesis that they resulted from random sampling from normally distributed populations were first determined for the average size 777 of the measurements, and the distributions of the observed β's were found to be impossible on this hypothesis. But this left an opening for confirmed believers in the universality of the normal distribution to assert that the impossibility arose from the samples being of various sizes. Accordingly an attempt was made to correct for this variation in size.

(3) The variation of the distribution curves of $\sqrt{\beta_1}$ and β_2 as depending on the size of the sample is of two kinds: (i) that which depends on the size of the curve or on its standard deviation. This difficulty could be got over by reducing each β to a common standard deviation, i.e. by multiplying it by the ratio of the standard deviation of a sample of 777 to its own sample S.D. This of course will modify the

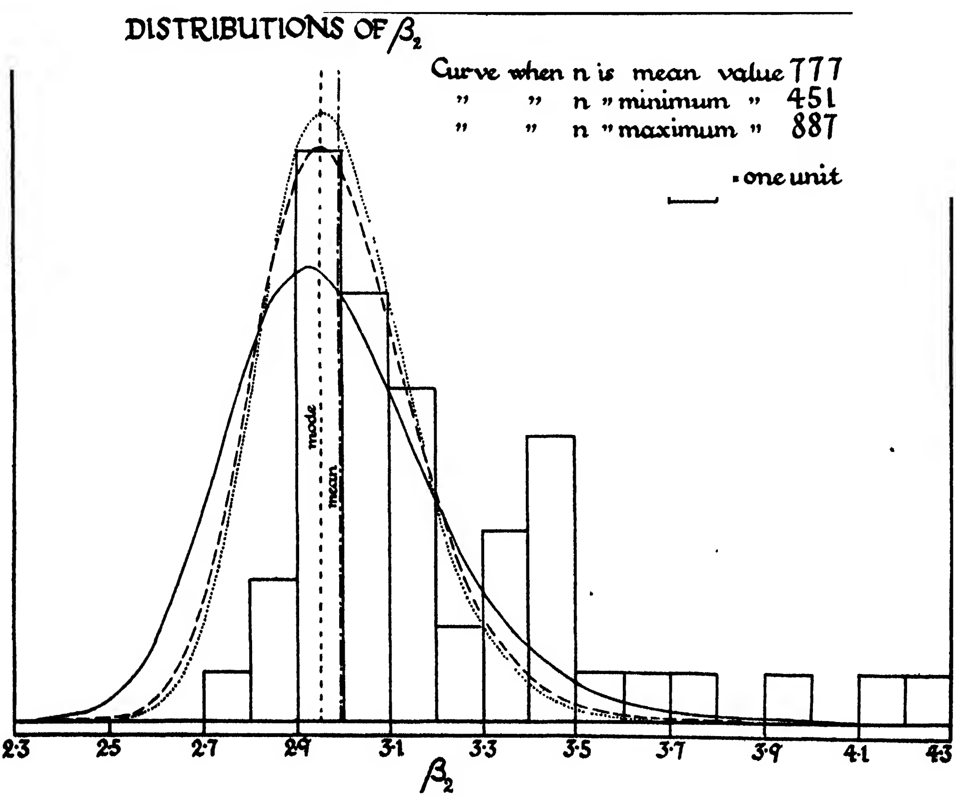


Diagram I.

ordinates of the curve, so that the y_0 must be multiplied by the inverse ratio to maintain the constant area which must in our case be equal to 50 units. (ii) Further, not only does the dimension of the distribution curve depend upon the size of the sample, but its shape does so also*. The problem accordingly arises as to how far after change of size in (i), the changes in shape will be of great or small importance.

Diagram I shows the histogram of the uncorrected observed values of β_2, and three curves of distribution of β_2 on the assumption that $n =$ the mean value 777, the minimum value 451 and the maximum value 887. The mean of the observed values is shown, but to avoid confusion only the mode and mean of the curve of mean value. It will be observed that there is considerable separation of the three curves of β_2 distribution, but it is clear that no curve in the region in which they lie would fit the observed values. Diagram II shows the data corrected by the method indicated, and the three curves reduced to the same standard deviation. The curves are not absolutely of the same shape, although they have the same S.D. and area; but they are so close together, and any curve lying between them, i.e. with a value of n between 451 and 887, will be so near to them, and all three so near to the mean curve for $n = 777$, that we may fairly take that curve as the theoretical distribution, which the corrected β_2's ought to follow.

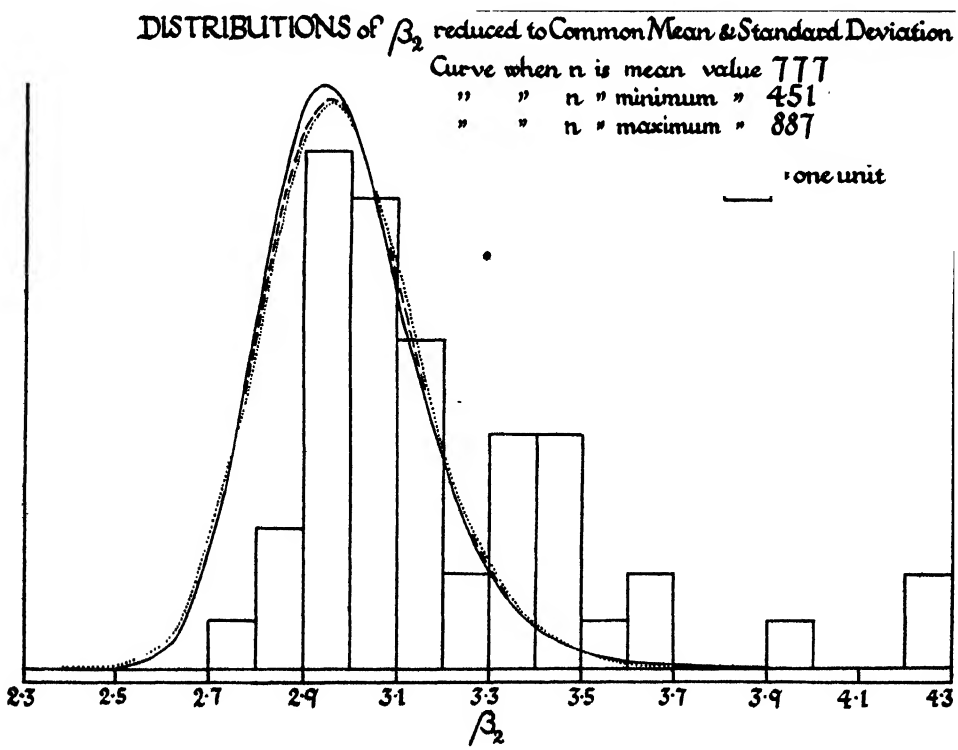


Diagram II.

* The B_1, B_2 of $\sqrt{\beta_1}$ and β_2 are functions of n, the size of the sample.

Even a cursory inspection of the diagram suffices to indicate that the β_2's are not a reasonable sample from this theoretical curve.

(4) Turning now to $\sqrt{\beta_1}$, its distribution follows a symmetrical curve and the character of this curve largely depends upon $B_2(\sqrt{\beta_1})$. This takes the values:

$$n = 451, \qquad n = 777, \qquad n = 887,$$

$$B_2(\sqrt{\beta_1}) = 3{\cdot}075{,}703 \qquad = 3{\cdot}044{,}926 \qquad = 3{\cdot}039505^*.$$

Thus we see that the distribution of $\sqrt{\beta_1}$ will closely follow a Type VII, but it will be very near to a normal distribution for all three cases, and accordingly it will be adequate to adjust the standard deviations to a common standard, namely that of $n = 777$. This reduction has been done and the modified histogram of the observed values of $\sqrt{\beta_1}$ fitted to the appropriate curve in Diagram III.

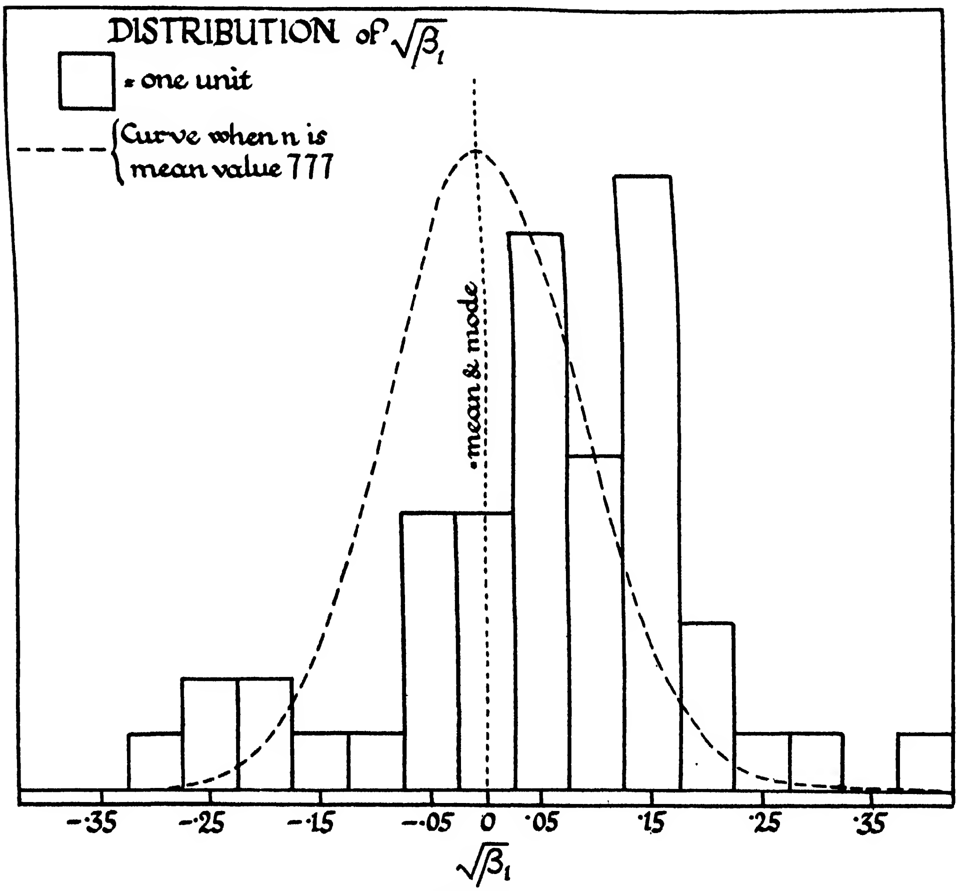


Diagram III.

It is clear from Diagrams II and III that it is impossible to look upon Dr Woo's 50 values each of $\sqrt{\beta_1}$ and β_2 as corresponding with 50 random samples

* See note on p. 47.

from material following the normal law of distribution. The conclusion is inevitably forced upon us that the normal curve is inadequate to describe effectively the distribution of measurements taken on single bones of the human skull. Such distributions show a defect in platykurtic and an excess in leptokurtic distributions from what would arise in samples from a normally distributed material. Furthermore the skewness shows a marked trend in the positive sense, that is to say the mean more frequently takes a higher than a less value than the mode; this is not consistent with random sampling from a normal population. It may be linked with the fact that variations in defect of the modal value must be largely limited in the case of small bones*. Having indicated the general results of these investigations, we may now point out the analysis by which they have been obtained.

(5) The values of $\sqrt{\beta_1}$ and β_2 for the 50 series of measurements with the numbers in each series are given in Table I. The first column gives the index

TABLE I. *Dr Woo's Values for* $\sqrt{\beta_1}$ *and* β_2.

Bone Measurement	Unadjusted $\sqrt{\beta_1}$		Adjusted $\sqrt{\beta_1}$		Unadjusted β_2		Adjusted β_2	
	Right	Left	Right	Left	Right	Left	Right	Left
F_4	+·176,535	+·136,190	+·188,324	+·145,286	3·386,011	3·459,867	3·40848	3·48556
F_5	+·158,138	+·128,279	+·168,883	+·136,996	3·206,534	3·013,689	3·21817	3·01395
P_2	+·037,597	−·004,648	+·037,043	−·004,579	2·980,196	3·359,173	2·98072	3·35076
P_3	+·028,811	−·042,832	+·030,874	−·045,899	2·982,362	3·012,825	2·98067	3·01305
P_4	−·094,144	−·269,368	−·091,774	−·262,588	3·413,815	3·179,417	3·39949	3·17328
O_7	+·236,230	+·106,875	+·246,001	+·112,653	3·077,681	3·056,409	3·08067	3·05845
O_8	+·398,165	+·268,491	+·418,280	+·282,055	3·623,313	2·953,096	3·64898	2·95070
O_9	+·087,327	+·160,650	+·091,739	+·168,766	3·189,819	3·424,293	3·19733	3·44163
T_1	+·093,172	+·099,974	+·098,327	+·105,506	2·954,450	3·121,385	2·95187	3·12654
T_2	−·006,726	+·042,595	−·007,053	+·044,662	2·857,046	2·983,474	2·85079	2·98235
T_3	+·149,535	+·152,579	+·158,268	+·161,490	3·044,128	3·119,839	3·04582	3·12527
T_4	+·042,170	−·155,341	+·040,777	−·150,210	2·997,727	3·446,918	2·99802	3·42241
T_5	+·121,357	+·106,378	+·128,802	+·112,904	2·930,373	2·958,346	2·92620	2·95569
T_6	+·060,248	+·049,088	+·062,746	+·051,123	3·017,157	3·131,634	3·01723	3·13418
T_7	−·031,082	−·059,960	−·032,311	−·062,331	2·850,937	3·082,440	2·84608	3·08456
Ml_1	+·009,545	+·002,617	+·009,786	+·002,683	3·571,235	4·185,434	3·58032	4·20456
Ml_2	−·290,128	−·277,429	−·279,004	−·266,792	4·267,608	3·954,612	4·20751	3·90941
Mx_1	−·206,880	−·199,216	−·195,342	−·188,106	3·148,294	3·273,137	3·14064	3·25864
Mx_2	−·077,979	+·012,537	−·063,731	+·010,246	3·413,535	3·197,307	3·34125	3·16376
Mx_3	+·051,303	+·044,662	+·039,195	+·034,122	2·919,089	2·996,785	2·94039	3·00003
Mx_4	+·246,201	+·257,146	+·206,540	+·215,722	3·746,382	3·485,451	3·62929	3·40983
S_2	+·145,442	+·130,073	+·139,959	+·125,169	3·337,501	3·078,842	3·32202	3·07540
S_3	−·065,375	+·067,052	−·068,992	+·070,762	3·316,481	2·930,254	3·33045	2·92659
S_5	+·039,711	+·135,358	+·038,294	+·130,527	2·760,055	3·018,663	2·77095	3·01805
S_6	+·112,832	+·134,687	+·115,054	+·137,339	2·868,398	2·994,103	2·86674	2·99383

* Cf. the positively skew curves obtained in the cases of age at scarlet fever attack, interest on securities, or period after marriage at which divorce proceedings are taken, etc., where there is a definite limit close to and below the modal value.

letter of the measurement, the second the number of measurements in the series, the third the observed value of $\sqrt{\beta_1}$ with its sign following that of μ_3, and the fourth gives the observed value of β_2.

The following formulae have been used in considering the distributions of $\sqrt{\beta_1}$ and β_2 as samples from a normal population. They presuppose the samples taken from a normal population to be of constant size, a condition not fulfilled by Dr Woo's material as explained above.

$$\text{Mean } \sqrt{\beta_1} = 0, \qquad \sigma_{\sqrt{\beta_1}} = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}},$$

$$B_1(\sqrt{\beta_1}) = 0, \qquad B_2(\sqrt{\beta_1}) = 3 + 36\frac{(n-7)(n^2+2n-5)}{(n-2)(n+5)(n+7)(n+9)} \quad \text{......(i)*},$$

$$\text{Mean } \beta_2 = 3\frac{n-1}{n+1}, \qquad \sigma_{\beta_2} = \frac{\sqrt{24}}{n+1}\sqrt{\frac{n(n-2)(n-3)}{(n+3)(n+5)}},$$

$$B_1(\beta_2) = \frac{216}{n}\frac{(n+3)(n+5)(n^2-5n+2)}{(n-3)(n-2)(n+7)^2(n+9)^2},$$

$$B_2(\beta_2) = 3 + \frac{36}{n}\frac{15n^6-36n^5-628n^4+982n^3+5777n^2-6402n+900}{(n-3)(n-2)(n+7)(n+9)(n+11)(n+13)} \quad \text{......(ii)*}.$$

For Dr Woo's material the minimum number of measurements in any series was 451, the maximum 887 and the average 777.

The above constants for these three cases are as follows:

	451		777		887	
	$\sqrt{\beta_1}$	β_2	$\sqrt{\beta_1}$	β_2	$\sqrt{\beta_1}$	β_2
Mean	0	2·98673	0	2·99229	0	2·99324
Standard Deviation	·11458	·22689	·08754	·17406	·08197	·16311
B_1	0	·449,323	0	·267,851	0	·235,714
B_2	3·075,703	4·102,966	3·044,926	2·662,507	3·039,505	3·583,784

all three distributions can be described by Type IV curves in the case of β_2 and by Type VII curves which are fairly close to the normal curve in the case of $\sqrt{\beta_1}$.

Dealing with $\sqrt{\beta_1}$ first the theoretical distribution for $n = 777$ is the curve

$$y = 4{\cdot}5826\left(1 + \frac{x^2}{1{\cdot}038{,}687^5}\right)^{-67{\cdot}276{,}031} \quad \text{.....................(iii)},$$

compared with the normal curve we have the following subrange frequencies:

* See *Biometrika*, Vol. XXII. p. 423.

Subrange Frequencies of Normal and Type VII Curves.

$\sqrt{\beta_1}$ (Central Value)	Normal Curve	Type VII Curve	Observed Values reduced
− ·40	·001	·001	—
− ·35	·004	·006	—
− ·30	·037	·040	1
− ·25	·212	·218	2
− ·20	·886	·873	2
− ·15	2·692	2·664	1
− ·10	5·956	5·924	1
− ·05	9·591	9·622	5
0	11·240	11·301	5
+ ·05	9·591	9·622	10
+ ·10	5·956	5·924	6
+ ·15	2·692	2·664	11
+ ·20	·886	·873	3
+ ·25	·212	·218	1
+ ·30	·037	·040	1
+ ·35	·004	·006	—
+ ·40	·001	·001	1
Actual Total = 50	49·998	49·997	50

It will be evident from the above values that for our present purpose Type VII curve is equivalent to a normal curve, and that accordingly we can legitimately pool our $\sqrt{\beta_1}$'s after the modification due to multiplying them by the factor ${}_m\sigma_{\beta_1}/{}_n\sigma_{\beta_1}$, and compare their distribution with the curve (iii). This is done graphically in Diagram III: see p. 48. The discrepancy even for a sample of 50 is marked. The difference of mean and standard deviation from the observed values are as follows:

	Theoretical Value	*Observed Value*
Mean:	0	·04897
Standard Deviation:	·08754	·14112

Only 14 out of the 50 $\sqrt{\beta_1}$'s are *negative* and the variation is greater than the theoretical value by more than 50%. If we apply the test of goodness of fit we find $\chi^2 = 100·058$ and $P <$ ·000,0005 for 11 groups. Remembering, however, that $\sqrt{\beta_1}$'s of the left and right sides are highly correlated, one might reach another limit by supposing we had only half the above frequencies, in which case χ^2 would equal 50·029, but we still find $P <$ ·000,0005. We thus see that the values of $\sqrt{\beta_1}$ are incompatible with random sampling from normal material.

In regard to the correlation of the $\sqrt{\beta_1}$'s for the homologous measurements left and right, we have the following constants, when we work with Dr Woo's uncorrected data.

	Left	*Right*
Mean $\sqrt{\beta_1}$:	·041,057	·056,880
Standard Deviation of $\sqrt{\beta_1}$:	·140,847	·140,951

Correlation of Right and Left $\sqrt{\beta_1}$'s $= \cdot 8474$.

The mean of all fifty uncorrected values $= \cdot 048{,}969$ and their standard deviation $= \cdot 141{,}121$.

This correlation in itself is a strong argument against the distributions being merely random samples from normal populations.

(6) We now turn to the distribution of the β_2's. Here we may determine in the first place the correlation between homologous β_2's. We find for the uncorrected values:

	Left	*Right*
Mean β_2:	3·216,696	3·194,405
Standard Deviation of β_2:	·304,413	·339,406

Correlation of Right and Left $\beta_2 = \cdot 6258$.

Now formula (ii) and the results in the table below it indicate that the mean β_2, whatever n may be, is less than 3, while the standard deviation for the shortest series is less than ·25. Hence it does not look even when the β_2's are adjusted as if we shall have a distribution corresponding to what may be anticipated in samples from normal populations. The correlation between the β_2's for the distributions of measurements on homologous bones confirms what has been observed with regard to the correlation of the $\sqrt{\beta_1}$'s, although the correlation coefficient is not so high, i.e. both emphasise that the skewness and kurtosis of an individual measurement are something peculiar to the bone on which the measurement is taken and are not due to the variations characteristic of random sampling.

In order to test the effect on β_2 of the variation in the sizes of the series, curves of Type IV were determined by the constants of the theoretical distributions for totals of 451, 777 and 887. The equations to these curves are as follows for the case of 50 samples:

(*a*) $n = 451$.

$$x = \cdot 712{,}387 \tan\theta, \qquad y = \cdot 2393855 \cos^{20\cdot 557{,}738}\theta\, e^{+16\cdot 399{,}738\theta}.$$

Origin at 2·357,179, Mode = 2·925,479,

and as already noted: Mean = 2·986,726.

(*b*) $n = 777$.

$$x = \cdot 687{,}983 \tan\theta, \qquad y = \cdot 0316555 \cos^{29\cdot 553{,}232}\theta\, e^{+23\cdot 048{,}097\theta}.$$

Origin at 2·416,795, Mode = 2·953,343,

and as already noted: Mean = 2·992,288.

(*c*) $n = 887$.

$$x = 683{,}293 \tan\theta, \qquad y = \cdot 01662195 \cos^{32\cdot 598{,}663}\theta\, e^{+25\cdot 354{,}120\theta}.$$

Origin at 2·427,065, Mode = 2·958,507,

and as already noted: Mean = 2·993,288.

The ordinates of these three curves were computed and they are plotted in Diagram I along with the original unadjusted values of β_2. The "mode" and "mean" are those corresponding to curve (*b*). It is clear that size of sample largely influences the form of the curve, and although the results are very improbable on the basis of any one of these curves, it is desirable to adjust the β_2's so that we may have a single curve for purposes of comparison. Now in the case of $\sqrt{\beta_1}$ all the corresponding three curves were so nearly normal that we had little hesitation in adjusting our $\sqrt{\beta_1}$'s by merely altering their values in the ratio of ${}_{777}\sigma_{\sqrt{\beta_1}}$ to ${}_{n}\sigma_{\sqrt{\beta_1}}$. We have attempted a similar method in the case of β_2, but *à priori* we cannot be equally confident of success, because the variation in *n* changes the shape as well as the size of the curve. Diagram II (p. 47) shows the distribution of the observed β_2's adjusted to a common mean and standard deviation. This is done in the following manner. In the place of ${}_{n}\beta_2$ we take

$$ {}_{777}\bar{\beta}_2 + ({}_{n}\beta_2 - {}_{n}\bar{\beta}_2)\, {}_{777}\sigma_{\beta_2} / {}_{n}\sigma_{\beta_2} $$

and it is this value we have termed "the adjusted value" of the observed ${}_{n}\beta_2$.

A similar treatment has been applied to the abscissae of the three curves (*a*), (*b*) and (*c*). This obviously does not change curve (*b*) at all, but it changes both (*a*) and (*c*). In the former case the abscissae are stretched in the ratio of 2·99229 to 2·98673, and in the latter case squeezed in the ratio 2·99229 to 2·99324. In both cases the ordinates or y_0's must be adjusted in the inverse ratio to retain the constant 50-units area. Diagram III (p. 48) indicates that while the three curves are not by this process brought into absolute agreement, they are sufficiently near to one another to justify our using the theoretical curve for $n = 777$ to describe the adjusted observed β_2's, on the supposition that the originals arose from sampling from normal populations. It is hardly needful to emphasise the badness of the fit!

The theoretical mean and standard deviation are 2·99229 and ·17406, while the observed are 3·19200 and ·31718. Thus the system of β_2's is more leptokurtic, and far more variable than is to be anticipated in sampling from normal populations.

The comparison of the observed and theoretical frequencies may be made from the table on p. 54.

The χ^2 for 10 groups is 266·869, which gives a probability of less than ·000,0005, and even if we suppose the β_2's from right and left sides *perfectly* correlated the χ^2 of 133·434 would also have a probability less than ·000,0005.

(7) The general conclusion to be drawn from Dr Woo's 50 $\sqrt{\beta_1}$'s and 50 β_2's seems to be that their distributions differ markedly from such as would arise from random sampling on the assumption that the distributions were each sampled from a normal population. We are accordingly forced to the important conclusion that the distributions of characters measured on the individual bones of the skull are not of normal type, but rather that the skewness and kurtosis of such distributions are peculiar to the individual measurement. It is possible that this can be accounted for by the failure of these measurements on individual cranial bones to be the product

Central Value	Adjusted Observed Value		Curve (b)	
2·45	—	1	·01	6·16
2·55	—		·18	
2·65	—		1·29	
2·75	1		4·68	
2·85	3		9·48	
2·95	11		11·87	
3·05	10		10·14	
3·15	7		6·50	
3·25	2		3·37	
3·35	5		1·51	
3·45	5		·61	
3·55	1	3	·23	0·31
3·65	2		·08	
3·75	—	3	·03	·0450
3·85	—		·01	
3·95	1		·0035	
4·05	—		·0015	
4·15	—		·0000	
4·25	2		·0000	
Totals	50		49·9950	

of a *large* number of contributory causes, for in such cases the proof of the normal law fails. We thus reach the suggestive conception that the simpler the organ measured, the less likely we are to meet with normal distributions. A measurement which depends upon growth from a few centres or even from a single centre of ossification is little likely to have its $\beta_1=0$ and $\beta_2=3$. But what is it that fixes the β_1 and β_2 in such a case? That is a problem of much interest and well worthy of study.

THE SAMPLING DISTRIBUTION OF THE THIRD MOMENT COEFFICIENT—AN EXPERIMENT.

By JOSEPH PEPPER, M.A., B.Sc.

1. The distributions of the mean and variance in samples from a normal population are known exactly. This, however, is not the case with the sampling distribution of the third moment coefficient

$$\nu_3 = \frac{1}{n} S(x - \bar{x})^3 \quad \text{.................................(1)},$$

where x is an individual observation in a sample whose mean is $\bar{x}$ and size is n. Although the equation to this sampling distribution is not yet known, we can, by actual sampling experiments, obtain valuable indications as to what form of distribution to expect.

2. In my paper on "Studies in the Theory of Sampling,"* I gave the mean and standard deviation of the distribution of ν_3 in sampling from any population. Their values in the case when the population is normal, with mean at the origin and standard deviation σ, may be written

$$_{\nu_3}M_1 = 0, \quad _{\nu_3}M_2 = \frac{6(n-1)(n-2)}{n^3}\sigma^6 \quad \text{...................(2)}.$$

The value of the third moment (and all the odd moments) of the distribution is zero, while the fourth semi-invariant has been given in the general case of sampling from any population by Dr R. A. Fisher†. These values of the moments for a normal population are:

$$_{\nu_3}M_3 = 0, \quad _{\nu_3}M_4 = 108n^{-6}(n-1)(n-2)(n^2+27n-70)\sigma^{12} \quad \text{.........(3)}.$$

This last value has also been obtained by C. C. Craig‡ (using semi-invariants) and P. R. Rider§.

3. From these values of the first four moments we derive

$$\beta_1 = \frac{M_3^2}{M_2^3} = 0,$$

$$\beta_2 = \frac{M_4}{M_2^2} = 3 + \frac{18(5n-12)}{(n-1)(n-2)} \quad \text{........................(4)}.$$

* *Biometrika*, Vol. XXI. Dec. 1929, pp. 235 and 238.

† "Moments and Product Moments of Sampling Distributions." *Proc. London Math. Soc.*, Ser. 2, Vol. XXX. Pt. 3 (1928), pp. 199—238.

‡ *Metron*, Vol. VII. 59.

§ "Moments of Moments." *Proc. National Academy of Sciences*, Vol. XV. No. 5, pp. 430—434, May 1929.

Thus, if we wish to attempt to fit the distribution of ν_3 in samples from a normal population by a Pearson curve, using only the values of its first four moments, the appropriate one would be Type VII, viz.:

$$y = y_0 \left(1 + \frac{x^2}{a^2}\right)^{-m} \qquad \text{.................................(5).}$$

4. All the odd moments of this curve about its mean, the origin, are zero, while its second and fourth moments, obtained by integration, are

$$\mu_2 = \frac{a^2}{2m-3}, \qquad \mu_4 = \frac{3a^4}{(2m-3)(2m-5)} \qquad \text{..................(6),}$$

giving

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3\left(1 + \frac{2}{2m-5}\right) \qquad \text{...........................(7).}$$

Further, if N is the area of the curve, we have

$$y_0 = \frac{N}{a\sqrt{\pi}} \frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})} \qquad \text{...............................(8).}$$

Equating the values of β_2 in (4) and (7) we get

$$m = \frac{5}{2} + \frac{(n-1)(n-2)}{6(5n-12)} \qquad \text{..............................(9).}$$

TABLE I.

Group	Positive	Negative	Group	Positive	Negative
0—50	42	49	1000—1050	4	3
50—100	42	31	1050—1100	5	2
100—150	42	36	1100—1150	4	2
150—200	26	21	1150—1200	1	7
200—250	26	30	1200—1250	2	4
250—300	28	15	1250—1300	1	1
300—350	15	17	1300—1350	1	4
350—400	19	12	1350—1400	1	4
400—450	15	15	1400—1450	1	0
450—500	11	2	1450—1500	3	1
500—550	9	16	1500—1550	0	0
550—600	7	8	1550—1600	0	0
600—650	4	3	1600—1650	1	1
650—700	7	8	1650—1700	1	1
700—750	11	9	1700—1750	2	1
750—800	7	4	1750—1800	2	0
800—850	7	15	1800—1850	0	0
850—900	5	3	1850—1900	2	0
900—950	4	2	1900—1950	1	0
950—1000	3	2	1950—2000	0	0

Outlying Frequencies.

Group	Positive	Group	Negative
2150—2200	1	2100—2150	3
2200—2250	1	2250—2300	1
3500—3550	1	2550—2600	1
		3500—3550	1

5. For the sampling experiment, samples of 10 from a normal population with mean at the origin and standard deviation 10 units, were already available, obtained by the use of Tippett's random sampling numbers.

From these samples, 700 values of the third moment coefficient ν_3 were computed. This was done by evaluating in the usual way, for each sample of 10,

$$\nu_1' = S(x), \quad \nu_2' = S(x^2), \quad \nu_3' = S(x^3),$$

and

$$\nu_3 = \nu_3' - 3\nu_1'\nu_2' + 2\nu_1'^3 \quad \text{.........................(10).}$$

The resulting frequency distribution of ν_3 is shown in Table I. The size of each group is taken to be 50 units for convenience, the plus column referring to the frequencies of ν_3 occurring as positive values and the minus column to negative values of ν_3.

6. For the determination of the constants of the Type VII curve, we have in the experiment, $n = 10$, $N = 700$ and $\sigma = 10$, so that from (9) and (2) we obtain

$$m = 2{\cdot}815{,}789, \quad {}_{\nu_3}M_2 = {\cdot}432 \times 10^6, \quad \sigma_{\nu_3} = 657{\cdot}3 \quad \text{...........(11).}$$

These values give, from equation (6),

$$a^2 = 1{\cdot}136{,}842 \times 10^6 \quad \text{..............................(12).}$$

We can now find y_0 in equation (8), which gives

$$\log y_0 = \bar{1}{\cdot}7277501, \quad y_0 = {\cdot}53426 \quad \text{....................(13).}$$

Thus, the curve which is to be fitted is

$$y = {\cdot}53426 \left\{1 + \frac{x^2}{(1{\cdot}136{,}842)\,10^6}\right\}^{-2{\cdot}815{,}789} \quad \text{.................(14).}$$

This curve, together with the observed frequency distribution of ν_3, are illustrated in Figure 1. The class interval in the histogram has been extended to 100 units so as to smooth out the irregularities of the observed frequencies in the class intervals of 50 units.

For the purpose of finding the moments of the observed distribution about its mean, it is convenient to take the values of ν_3 at $25a\,(a = 1, 3, 5, \ldots)$ as $+1$, $+3$, $+5$ and so on, and similarly for the negative values. The values of the first four moments, in terms of the units of the sampled population, are

$$\begin{aligned}
{}_{\nu_3}M_1 &= 0{\cdot}3457 \times 25 = 8{\cdot}64,\\
{}_{\nu_3}M_2 &= 609{\cdot}87 \times 25^2, \quad \sqrt{{}_{\nu_3}M_2} = 617{\cdot}4,\\
{}_{\nu_3}M_3 &= -1486{\cdot}074 \times 25^3,\\
{}_{\nu_3}M_4 &= 2793105{\cdot}4 \times 25^4.
\end{aligned}$$

The standard error of ${}_{\nu_3}M_1$ is $\dfrac{657{\cdot}3}{\sqrt{700}} = 24{\cdot}844$ and the standard error of $\sqrt{{}_{\nu_3}M_2}$ will be given approximately by the formula

$$\sigma_\sigma = \frac{\sigma}{2}\sqrt{\frac{\beta_2 - 1}{N}},$$

where in our case, $\sigma = 657{\cdot}3$, $\beta_2 = 12{\cdot}5$, $N = 700$, so that the standard error of $\sqrt{{}_{\nu_3}M_2}$ is $42{\cdot}12$.

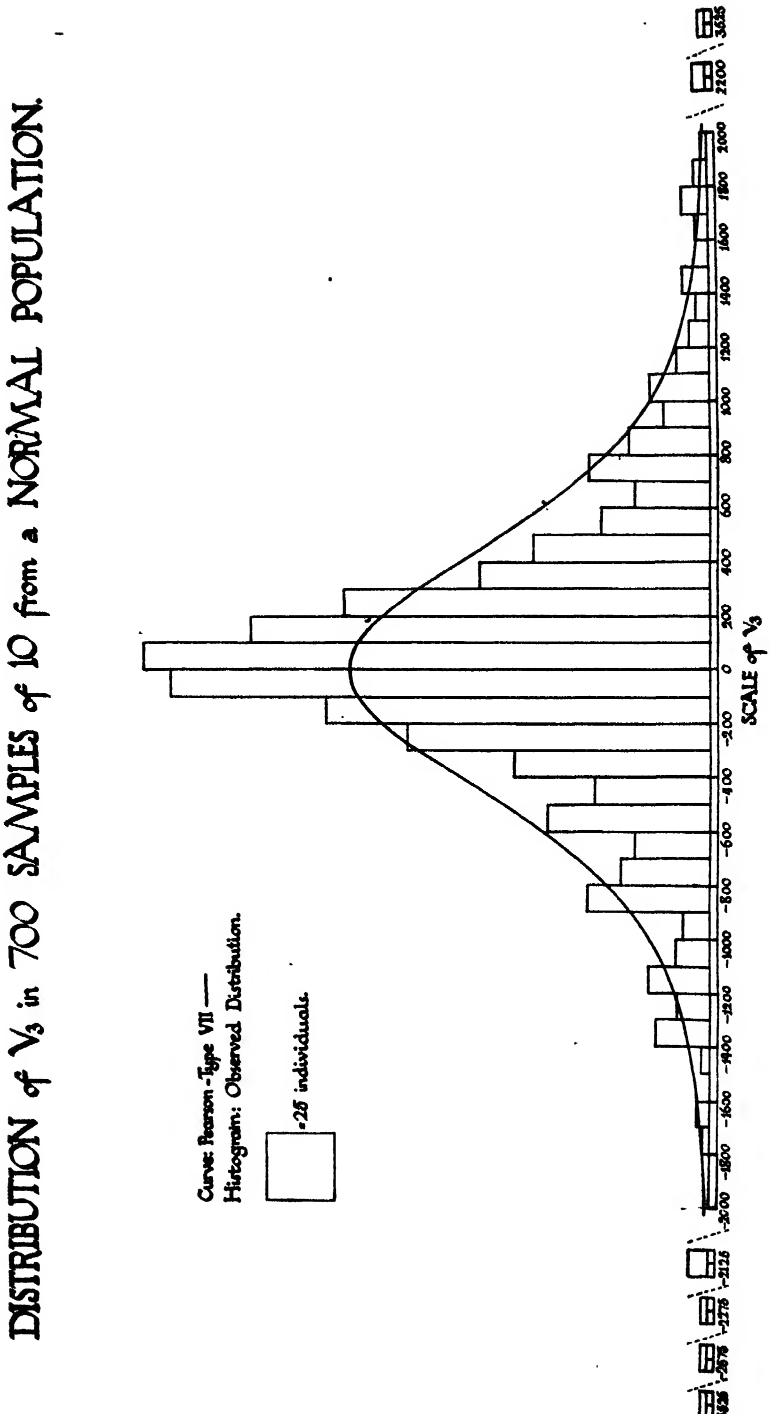

DISTRIBUTION of V_3 in 700 SAMPLES of 10 from a NORMAL POPULATION.
Curve: Pearson-Type VII —
Histogram: Observed Distribution.
=25 individuals.
SCALE of V_3
-3626
-2675
-2275
-2125
-2000
-1800
-1600
-1400
-1200
-1000
-800
-600
-400
-200
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
3625

Fig. 1.

Thus the mean and standard deviation of the observed distribution of ν_3 differ from their theoretical values by less than their standard errors.

Further, the β_1 and β_2 of the observed distribution are given by

$$\beta_1 = \frac{M_3^2}{M_2^3} = \cdot 010,$$

$$\beta_2 = \frac{M_4}{M_2^2} = 7 \cdot 51.$$

The theoretical values of β_1 and β_2 are 0 and 12·5 respectively, so that the observed β_2 is rather small.

7. To determine the goodness of fit of the Type VII curve to the distribution of ν_3, the observed frequencies n_s' in Figure 1 and the corresponding frequencies n_s of the curve, calculated from the formula

$$n_s = \frac{h}{24} \{y_{s-1} + 22 y_s + y_{s+1}\}$$

are shown in Table II. The frequencies are divided into 30 groups; those corresponding to positive and negative values of ν_3 shown in rows (1) and (3) respectively.

TABLE II.

Group	0—1	1—2	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12	12—13	13—14	Over 14
(1) n_s'	84	68	54	34	26	16	11	18	12	7	9	5	3	2	16
(2) n_s	52·9	50·0	46·0	40·0	33·5	27·9	22·0	17·0	13·5	10·0	8·0	6·0	4·5	3·5	15·2
(3) n_s'	80	57	45	29	17	24	11	13	18	4	5	9	5	8	10

It will be seen that the frequencies n_s' in the group (0—1) differ considerably from the corresponding n_s, and in fact contribute about one-third of the value of χ^2. This is given by

$$\chi^2 = \frac{S(n_s - n_s')^2}{n_s} = 90 \cdot 24,$$

yielding a value $P = 1 \cdot 746 \times 10^{-8}$.

Accordingly, the goodness of fit of the Pearson Type VII curve

$$y = y_0 \left(1 + \frac{x^2}{a^2}\right)^{-}$$

to the distribution of the third moment coefficient in samples from a normal population is unsatisfactory, so that it appears, in this case, that the first four moments do not furnish sufficient data for the determination of the best fitting curve.

8. For any explanation of the bad fit of the Type VII curve, it would be desirable to compare the values of the higher moments of the distribution of the third moment with the experimental values. A method for finding the sixth and

eighth moments was given by the work of Dr R. A. Fisher*. The quantity γ, whose 2nd, 4th and 6th moments are given in this paper, is equivalent to $\sqrt{n(n-1)\beta_1}/(n-2)$ in samples. Further, Dr Fisher finds that "the moment of the distribution of γ is derivable by multiplying by

$$\frac{(n-1)^r}{(n-1)(n+1)\ldots(n+2r-3)\kappa_2{}^r}$$

the corresponding moment of the distribution of k_3†."

Thus, from the value of $\mu_6(\gamma)$ it is found that

$$\mu_6(\nu_3) = 3240n^{-9}(n-1)(n-2)(n^4 + 84n^3 + 2695n^2 - 15168n + 20020)\,\sigma^{18}$$

and therefore

$$\beta_4(\nu_3) = 15\,(n^4 + 84n^3 + 2695n^2 - 15168n + 20020)/(n-1)^2(n-2)^2.$$

By extending the work of Dr Fisher for the case of the eighth moment I found that

$$\begin{aligned}\mu_8(\gamma) = 136080n^4(n-1)^4(n^6 + 171n^5 + 13893n^4 + 580401n^3 - 5131014n^2 \\ + 14132268n - 12932920) \\ \div (n-2)^7(n+1)(n+3)\ldots(n+21).\end{aligned}$$

This value has been checked by Dr Fisher who also evaluated it by his combinatorial method. From it, I obtain

$$\beta_6(\nu_3) = 105P_6/(n-1)^3(n-2)^3,$$

where P_6 is the sixth degree polynomial in the above expression for $\mu_8(\gamma)$.

9. Table III below gives the theoretical and experimental values of the β_2, β_4 and β_6 of the distribution of ν_3.

TABLE III.

	Theory	Experiment
β_2	12·5	7·51
β_4	670·83	139·04
β_6	99225	3702

Although we are ignorant of the values of the standard errors of β_4 and β_6, yet the divergencies shown in Table III are so large that at first sight it appeared there must be some fundamental error in the sampling. However, it must be remembered that these higher theoretical betas of the distribution take account of values of ν_3, no matter how large, while in the experiment, a limited number of samples, namely 700, only can be used. The addition of one value of ν_3 at a sufficient distance from the origin would greatly increase the values of β_4 or β_6.

* $k_3 = \dfrac{n^2}{(n-1)(n-2)}\nu_3$ in samples.

† *Proc. Roy. Soc.* Ser. A, Vol. 130, No. A, 812, p. 16.

We may further note that β_4 and β_6 in the case of the Type VII curve are given by

$$\beta_4 = \frac{5\beta_2^2}{6-\beta_2}, \qquad \beta_6 = \frac{7\beta_2\beta_4}{9-2\beta_2}$$

which both take impossible negative values when $\beta_2 = 12\cdot5$ as given by the theory*.

10. It was next decided, as the sample values of ν_3 were available, to evaluate the $\sqrt{\beta_1} = \nu_3\nu_2^{-\frac{3}{2}}$ for each sample, and study the distribution of this quantity, analogous to the γ of Dr Fisher's paper. Theoretically, it is distributed symmetrically and we further know the first eight moments, that is, the β_2, β_4 and β_6 of the distribution. It is independent of the units of measurement, and the introduction of the term $\nu_2^{-\frac{3}{2}}$ has the effect of bringing the outlying values of ν_3 nearer the origin.

The frequency distribution of $\sqrt{\beta_1}$ is shown in Table IV, the size of the groups being taken equal to 0·1, as most convenient. The mean, standard deviation, β_2, β_4 and β_6 of the distribution are as follows:

$$\begin{aligned} &\text{Mean } \sqrt{\beta_1} = 0\cdot0289, \quad \sigma^2_{\sqrt{\beta_1}} = 0\cdot347067, \\ &\sigma_{\sqrt{\beta_1}} \quad\quad = 0\cdot5891, \quad \beta_2 \ = 3\cdot3370, \\ &\beta_4 \quad\quad\quad = 18\cdot89, \quad \beta_6 \ = 139\cdot03. \end{aligned}$$

The theoretical value of β_2 is given by

$$\beta_2 = \frac{3(n+1)(n+3)(n^2+27n-70)}{(n-2)(n+5)(n+7)(n+9)}$$

which in the case of $n = 10$ gives $\beta_2 = 3\cdot3204$, so that it was again decided to fit a Type VII curve to the distribution. If the equation is, as before,

$$y = y_0\left(1 + \frac{x^2}{a^2}\right)^{-m},$$

TABLE IV.

Group	0·0—0·1	0·1—0·2	0·2—0·3	0·3—0·4	0·4—0·5	0·5—0·6	0·6—0·7	0·7—0·8	0·8—0·9	0·9—1·0	1·0—1·1	1·1—1·2	1·2—1·3	1·3—1·4	1·4—1·5	1·5—1·6	1·6—1·7	1·7—1·8	1·8—1·9
+ve	43	58	52	39	30	30	24	21	17	15	11	9	7	1	2	3	1	1	1
−ve	49	47	44	36	35	30	22	22	15	12	4	4	4	1	2	2	2	0	4

the values of y_0, a^2 and m are given in section 4, in terms of the first four moments. We thus obtain the Type VII curve to be fitted as

$$y = 499\cdot211\left(1 + \frac{x^2}{6\cdot956522}\right)^{-11.862319}$$

* It should be noted that for the Type VII curve which has been used for fitting the $\sqrt{\beta_1}$ distribution, the moments also become eventually impossible. This occurs with $\beta_{22} = 23\beta_2\beta_{20}/(33-10\beta_2)$ for the denominator becomes negative, since $\beta_2 = 3\cdot3204$.

This curve, together with the histogram in which the size of the groups has been extended to 0·2 to smooth frequency irregularities, is illustrated in Figure 2. As it is seen, the fit is remarkably good, in great contrast to the case of the ν_2 distribution. In finding the value of the goodness of fit, 26 groups have been used, the

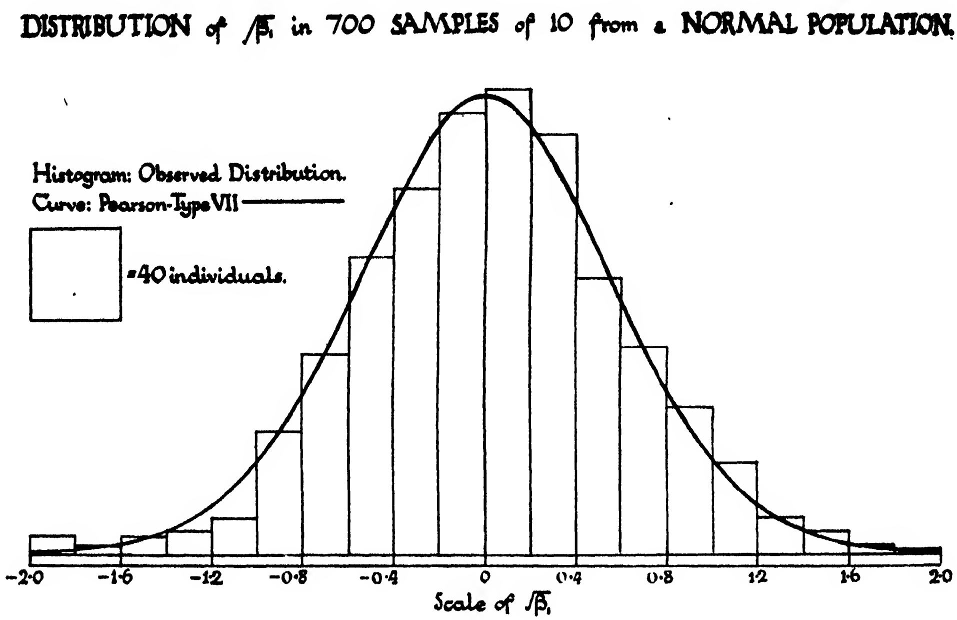


Fig. 2.

last 7 groups on each side of the origin being clubbed together. The values of the areas n_s of the Type VII curve, corresponding to the groups in Table IV, were calculated from the ordinates by the formula

$$n_s = \frac{h}{24}(y_{s-1} + 22y_s + y_{s+1})$$

and are as follows:

Group	·05	·15	·25	·35	·45	·55	·65	·75	·85	·95	1·05	1·15	Tail
n_s	49·7	48·0	44·9	40·6	35·5	30·1	24·8	19·9	15·5	11·8	8·7	6·3	14·2

The mid-points of the groups have been given, and the last group gives the area in the tail containing the last seven groups. Evaluating in the usual way $\Sigma \frac{(n_s - n_s')^2}{n_s}$ for the 26 groups, I obtained $\chi^2 = 12·70$, giving the very high value of $P = 0·98$. This excellent fit is to be expected when one compares the experimental, theoretical and Type VII values of the moments shown in Table V. The values for the normal curve have been added for comparison.

TABLE V.

	Experiment	Theory	Type VII	Normal
Mean ...	0·0289	0	0	0
St. Dev.	0·5891	0·5794	0·5794	0·5794
β_2 ...	3·3370	3·3204	3·3204	3·0
β_4 ...	18·89	18·99	20·5722	15·0
β_6 ...	139·03	148·954	202·677	105·0

The standard error of the mean is $\cdot 5794/\sqrt{700} = \cdot 0219$ and that of the standard deviation is $\frac{\cdot 5794}{2}\sqrt{\frac{3 \cdot 3204 - 1}{700}} = \cdot 0167$, so that the experimental values differ from their actual values by approximately their standard errors. The standard errors of β_2, β_4 and β_6 are given approximately by the formulae

$$\sigma^2_{\beta_2} = \frac{1}{N}(\beta_6 - 4\beta_4\beta_2 + 4\beta_2^3 - \beta_2^2),$$

$$\sigma^2_{\beta_4} = \frac{1}{N}(\beta_{10} - 6\beta_6\beta_4 + 9\beta_4^2\beta_2 - 4\beta_4^2),$$

$$\sigma^2_{\beta_6} = \frac{1}{N}(\beta_{14} - 8\beta_8\beta_6 + 16\beta_6^2\beta_2 - 9\beta_6^2).$$

We must therefore make some estimate of the values of β_8, β_{10} and β_{14}. As a guide, we may consider their normal values, which are known, and the Type VII values given by

$$\beta_8 = \frac{3\beta_2\beta_6}{4 - \beta_2}, \quad \beta_{10} = \frac{33\beta_2^2\beta_6}{(15 - 4\beta_2)(4 - \beta_2)}, \quad \beta_{14} = \frac{65\beta_2^2\beta_{10}}{(7 - 2\beta_2)(18 - 5\beta_2)}.$$

These are given in Table VI.

TABLE VI.

	Normal	Type VII
β_8	945	2971
β_{10}	10395	63142
β_{14}	2027025	90109172

The estimates chosen were the mean value 1958 for β_8, 15,000 for β_{10} and the normal value of β_{14}. The standard errors are, on substitution of these values,

$$\sigma_{\beta_2} = \cdot 2143, \qquad \sigma_{\beta_4} = 3 \cdot 243, \qquad \sigma_{\beta_6} = 31.$$

Therefore it appears that experiment, Type VII, or even the normal values are in good accordance with theory.

11. The normal curve having the same mean and standard deviation as the theory was also fitted. The curve is

$$y = \frac{700}{\cdot 5794\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\cdot 5794}\right)^2}.$$

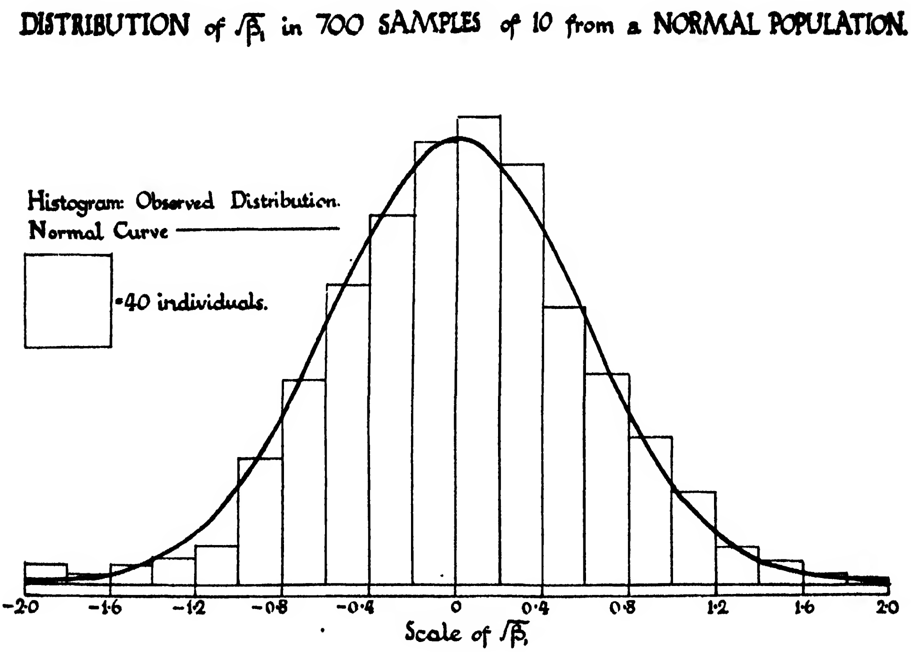


Fig. 3.

This curve with the histogram is shown in Figure 3. The values of the areas n_s of the normal curve corresponding to the same 26 groups are given below:

Group	·05	·15	·25	·35	·45	·55	·65	·75	·85	·95	1·05	1·15	Tail
n_s	47·96	46·56	43·87	40·13	35·63	30·71	25·70	20·87	16·46	12·59	9·36	6·75	13·41

The value of χ^2 in this case is 13·53 yielding $P = 0{\cdot}97$, again an exceedingly high value.

Conclusion. The satisfactory results given by the distribution of $\sqrt{\beta_1} = \nu_3\nu_2^{-\frac{3}{2}}$ in samples from a normal population indicate that it is better to work with this ratio than the third moment coefficient ν_3, whose distribution is so much more irregular. As the values of all the betas of the distribution of $\sqrt{\beta_1}$ tend to the normal values as the size of the sample increases, and the normal curve fits practically as well as the Type VII curve to the distribution, it would appear desirable in practice to take the normal curve as representing the distribution of $\sqrt{\beta_1}$ in samples, failing the discovery of the actual curve.

FURTHER CONTRIBUTIONS TO THE SAMPLING PROBLEM.

By N. St. GEORGESCU, D.Stat. Paris, Lecturer in the Bucharest School of Statistics, Board of Census, Bucharest.

1. Many important researches in statistics have lately dealt with the moments and product moments of sample distributions. Although, except in the case of the mean and the standard deviation, no law of distribution has been found for any higher moment or semi-invariant, the general problem has been elucidated in a great many points.

A few years ago C. C. Craig, in a very interesting paper, indicated several procedures for calculating the product moments and the semi-invariants of sampling moments. At the same time he gave a synthesis of the most important results already obtained, by previous authors, on the sampling problem. That is why, for bibliographic information, the reader is advised to refer to Craig's paper*.

Later, R. A. Fisher gave a practical procedure, based on combinatory analysis, for calculating the cumulative moments of the sampling cumulative moments†. Still more recently, the same author found out the first few moments of many important measures of the departure from normality‡.

As an illustration of the use of such results I may quote a paper in which E. S. Pearson has calculated two tables of probabilities for the distributions of β_1 and β_2 of large samples§. As one may see in the latter author's note the difference between the values of the moments used in these tables—calculated with approximation—and the true values can be neglected for most practical purposes‖.

The sampling problem has been handled in two different ways. The first considers the case of a sample of any size, the second provides formulae available for large samples only; but both these ways present difficulties.

Most methods given—up to the present—for dealing with a sample of any size do not enable us to reach a general expression for the sampling distribution, the procedure of calculation being entirely different in each case. As for the case of large samples, although general formulae may be worked out, they are only an approximation with respect to the powers of $1/N$, and if a better value is required

* C. C. Craig. "An Application of Thiele's Semi-invariants to the Sampling Problem." *Metron*, Vol. VII. (4), pp. 2—74 (1928).

† R. A. Fisher. "Moments and Product Moments of Sampling Distributions." *Proc. Lond. Math. Soc.*, Series 2, Vol. XXX. pp. 199—238 (1928).

‡ R. A. Fisher. "The Moments of the Distribution for Normal Samples of Measures of Departure from Normality." *Proc. Roy. Soc.* A, Vol. 130, pp. 16–28 (1930).

§ E. S. Pearson. "A Further Development of Tests for Normality." *Biometrika*, Vol. XXII. p. 239.

‖ E. S. Pearson. "Note on Tests for Normality." *Biometrika*, Vol. XXII. p. 423.

further calculations are necessary. Therefore efforts are being made to find out general formulae by one or other of the two methods mentioned above; nevertheless for practical purposes the case of large samples is much the more useful.

In this paper I shall introduce new functions in connection with the distributions of random variables and offer a new method—based on these functions—which will allow us to obtain either exact formulae for small samples, or approximated results for large samples. Applications will be made to the case of a normal population.

Before commencing to show the results I have obtained, I must not fail to acknowledge my gratitude to Professor Karl Pearson, by the aid of whose assistance and criticism these results were carried to their present stage.

Statistical Differentials.

2. Let us consider a parent population of which the distribution is the following:

Values	$x_1, x_2, x_3, \ldots x_t$,
Probabilities	$p_1, p_2, p_3, \ldots p_t$,

and any sample of N from that population:

Values	$x_1, x_2, x_3, \ldots x_t$,
Relative frequencies	$\varpi_1, \varpi_2, \varpi_3, \ldots \varpi_t$.

Let us design by $\delta\varpi_i$ the difference between ϖ_i and its mean value p_i:

$$\varpi_i = p_i + \delta\varpi_i.$$

The characteristic function of the distribution of the $\delta\varpi_i$'s will be*

$$f(u_1, u_2, \ldots u_t) = e^{-\Sigma p_i u_i} (\Sigma p_i e^{\frac{u_i}{N}})^N \qquad (1);$$

therefore in any product moment of the $\delta\varpi_i$'s the lowest power of $1/N$ is $I(k+1)$, where k is the order of the product moment considered, and $I(\alpha)$ denotes the greatest integer contained by α.

Now let us consider any function of the sample, i.e. of $\varpi_1, \varpi_2, \varpi_3, \ldots \varpi_t$, the x's being constant when the parent population does not change. Let $F(\varpi_1, \varpi_2, \ldots \varpi_t)$ be such a function; the same function, but with respect to the parent population, will be $F(p_1, p_2, \ldots p_t)$.

Between these two functions we have the following relation:

$$\begin{aligned}\delta F(\varpi_1, \varpi_2, \ldots \varpi_t) &= F(\varpi_1, \varpi_2, \varpi_3, \ldots \varpi_t) - F(p_1, p_2, p_3, \ldots p_t)\\ &= \Sigma_i \delta\varpi_i \frac{\partial F(p_1, p_2, \ldots p_t)}{\partial p_i}\\ &+ \frac{1}{2!} \Sigma_{i,j} \delta\varpi_i \delta\varpi_j \frac{\partial^2 F(p_1, p_2, \ldots p_t)}{\partial p_i \partial p_j}\\ &+ \frac{1}{3!} \Sigma_{i,j,k} \delta\varpi_i \delta\varpi_j \delta\varpi_k \frac{\partial^3 F(p_1, p_2, \ldots p_t)}{\partial p_i \partial p_j \partial p_k}\\ &+ \ldots \qquad (2).\end{aligned}$$

* See Note I, at the end.

We observe that in order to find out the moments of $\delta F(\varpi_1, \varpi_2, \varpi_3, \dots \varpi_t)$ with a certain degree of approximation, owing to the remark made above concerning the moments of $\delta\varpi_i$, we reach the first approximation in $1/N$ by taking the first term only of the right-hand side of the relation (2). That approximation would be rendered closer by taking more terms in the same relation.

Let $F'(f_1', f_2', f_3', \dots f_r')$ be a composite function of the sample, the corresponding function of the parent population being $F(f_1, f_2, f_3, \dots f_r)$. Supposing that we want an approximation of the second order, we shall have

$$\delta_2 f_k' = \Sigma_i \delta\varpi_i \frac{\partial f_k}{\partial p_i} + \frac{1}{2!}\Sigma_{i,j}\delta\varpi_i\delta\varpi_j \frac{\partial^2 f_k}{\partial p_i \partial p_j};$$

hence

$$\delta_2 F' = \Sigma_i \delta_2 f_i' \frac{\partial F}{\partial f_i} + \frac{1}{2!}\Sigma_{i,j}\delta_2 f_i' \delta_2 f_j' \frac{\partial^2 F}{\partial f_i \partial f_j},$$

where on the right-hand side we must neglect, of course, the terms of the third and fourth order in $\delta\varpi_i$; thus

$$\delta_2 F' = \Sigma_i \delta_2 f_i' \frac{\partial F}{\partial f_i} + \frac{1}{2!}\Sigma_{i,j}\delta_1 f_i' \delta_1 f_j' \frac{\partial^2 F}{\partial f_i \partial f_j} \quad \dots\dots\dots\dots\dots\dots(3).$$

In the same way

$$\delta_1 F' = \Sigma_i \delta_1 f_i' \frac{\partial F}{\partial f_i} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(4),$$

$$\begin{aligned}\delta_3 F' = {} & \Sigma_i \delta_3 f_i' \frac{\partial F}{\partial f_i} \\ & + \frac{1}{2!}\Sigma_{i,j}(\delta_2 f_i' \delta_1 f_j' + \delta_1 f_i' \delta_2 f_j' - \delta_1 f_i' \delta_1 f_j') \frac{\partial^2 F}{\partial f_i \partial f_j} \\ & + \frac{1}{3!}\Sigma_{i,j,k}\delta_1 f_i' \delta_1 f_j' \delta_1 f_k' \frac{\partial^3 F}{\partial f_i \partial f_j \partial f_k} \quad \dots\dots\dots\dots\dots\dots(5),\end{aligned}$$

$$\begin{aligned}\delta_4 F' = {} & \Sigma_i \delta_4 f_i' \frac{\partial F}{\partial f_i} \\ & + \frac{1}{2!}\Sigma_{i,j}(\delta_3 f_i' \delta_1 f_j' + \delta_2 f_i' \delta_2 f_j' + \delta_1 f_i' \delta_3 f_j' - \delta_2 f_i' \delta_1 f_j' - \delta_1 f_i' \delta_2 f_j') \frac{\partial^2 F}{\partial f_i \partial f_j} \\ & + \frac{1}{3!}\Sigma_{i,j,k}(\delta_2 f_i' \delta_1 f_j' \delta_1 f_k' + \delta_1 f_i' \delta_2 f_j' \delta_1 f_k' + \delta_1 f_i' \delta_1 f_j' \delta_2 f_k' \\ & \qquad - 2\delta_1 f_i' \delta_1 f_j' \delta_1 f_k') \frac{\partial^3 F}{\partial f_i \partial f_j \partial f_k} \\ & + \frac{1}{4!}\Sigma_{i,j,k,l}\delta_1 f_i' \delta_1 f_j' \delta_1 f_k' \delta_1 f_l' \frac{\partial^4 F}{\partial f_i \partial f_j \partial f_k \partial f_l} \quad \dots\dots\dots\dots\dots\dots(6).\end{aligned}$$

These formulae are very useful and similar ones—for higher degrees of approximation—may be easily obtained. They enable us to make a connection between mathematical and statistical differentials. We shall make a first application of the formula (4) to the sampling semi-invariants.

Approximate Formulae for the Product Moments of the Sampling Semi-invariants.

3. Let

$$e^{\phi(t)} = e^{s_1\frac{t}{1!} + s_2\frac{t^2}{2!} + s_3\frac{t^3}{3!} + \dots + s_p\frac{t^p}{p!} + \dots}$$

$$= 1 + m_1\frac{t}{1!} + m_2\frac{t^2}{2!} + m_3\frac{t^3}{3!} + \dots + m_q\frac{t^q}{q!} + \dots \quad \dots\dots\dots(7)$$

be the characteristic function of the parent population, the same function for the sample being

$$e^{s_1' \frac{t}{1!} + s_2' \frac{t^2}{2!} + \dots + s_p' \frac{t^p}{p!} + \dots} = 1 + m_1' \frac{t}{1!} + m_2' \frac{t^2}{2!} + \dots + m_q' \frac{t^q}{q!} + \dots \quad (8).$$

(In these two relations the notations are those usually adopted; i.e. the s's stand for *Thiele's* semi-invariants, and m for the moments about a fixed origin.)

Let us apply the formula (4) to the relation (8); we shall have

$$\frac{t}{1!} \delta_1 s_1' \frac{t^2}{2!} \delta_1 s_2' + \dots + \frac{t^p}{p!} \delta_1 s_p' + \dots = \left(\frac{t}{1!} \delta m_1' + \frac{t^2}{2!} \delta m_2' + \dots + \frac{t^q}{q!} \delta m_q' + \dots\right) e^{-\phi(t)} \quad \dots\dots\dots(9),$$

and if we put

$$e^{-s_1 \frac{t}{1!} - s_2 \frac{t^2}{2!} - s_3 \frac{t^3}{3!} + \dots} = 1 + M_1 \frac{t}{1!} + M_2 \frac{t^2}{2!} + \dots + M_q \frac{t^q}{q!} + \dots \quad \dots(10),$$

the relation (9) will give

$$\delta_1 s_p' = \delta m_p' + \binom{p}{1} M_1 \delta m'_{p-1} + \dots + \binom{p}{i} M_i \delta m'_{p-i} + \dots + \binom{p}{1} M_{p-1} \delta m_1' \quad \dots\dots\dots(A),$$

where $$M_p = \mathbf{S} \frac{(-1)^\rho}{\rho_1! \rho_2! \rho_3! \dots} \times \frac{p!}{(r_1!)^{\rho_1} (r_2!)^{\rho_2} \dots} s_{r_1}^{\rho_1} s_{r_2}^{\rho_2} s_{r_3}^{\rho_3} \dots \quad \dots\dots\dots\dots(B),$$

with $$\Sigma r_i \rho_i = p, \quad \Sigma \rho_i = \rho.$$

It is worth noticing that m' being a linear function of ϖ_i we have $\delta m_p' = \delta_k m_p'$.

4. The relation (A) enables us to calculate the first approximation of the product moment of any two semi-invariants. By using that formula we have

$$\{\delta_1 s_a' \delta_1 s_b'\} = \sum_{1}^{a}{}_i \sum_{1}^{b}{}_j \binom{a}{i} \binom{b}{j} M_{a-i} M_{b-j} \{\delta m_i' \delta m_j'\} \quad \dots\dots\dots\dots(11),$$

and since $$\{\delta m_i' \delta m_j'\} = \frac{1}{N} (m_{i+j} - m_i m_j),$$

the relation (11) becomes

$$\{\delta_1 s_a' \delta_1 s_b'\} = \frac{1}{N} \sum_{1}^{a}{}_i \sum_{1}^{b}{}_j \binom{a}{i} \binom{b}{j} M_{a-i} M_{b-j} m_{i+j} - \frac{1}{N} \sum_{1}^{a}{}_i \binom{a}{i} M_{a-i} m_i \sum_{1}^{b}{}_j \binom{b}{j} M_{b-j} m_j \quad \dots\dots\dots(12).$$

Now let us return to the relation (7), which by using (10) may be written down as follows:

$$\left(1 + m_1 \frac{t}{1!} + m_2 \frac{t^2}{2!} + \dots + m_p \frac{t^p}{p!} + \dots\right) \left(1 + M_1 \frac{t}{1!} + M_2 \frac{t^2}{2!} + \dots + M_q \frac{t^q}{q!} + \dots\right) = 1.$$

The comparison of the coefficients of the same powers of t gives

$$M_1 + m_1 = 0,$$
$$M_2 + 2M_1 m_1 + m_2 = 0,$$

$$M_\beta + \binom{\beta}{1} M_{\beta-1} m_1 + \dots + \binom{\beta}{j} M_{\beta-j} m_j + \dots + m_\beta = 0,$$

or $$\sum_{0}^{\beta}{}_j \binom{\beta}{j} M_{\beta-j} m_j = 0 \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(C).$$

That relation enables us to write down the following formulae:

$$\{\delta_1 s_a' \, \delta_1 s_b'\} = \frac{1}{N}\left[\sum_1^a{}_i \sum_1^b{}_j \binom{a}{i}\binom{b}{j} M_{a-i} M_{b-j} m_{i+j} - M_a M_b\right] \text{.........(13).}$$

This is a very convenient formula, especially for practical purposes, the expression which gives M_p being similar to that which gives m_p as a function of the semi-invariants, except for some of the signs. But for the result we are seeking we shall proceed to change its form. For that purpose let us consider the series

$$e^{-\phi(t)} \frac{d^i}{dt^i}[e^{\phi(t)}] = N_0^{(i)} + N_1^{(i)} \frac{t}{1!} + N_2^{(i)} \frac{t^2}{2!} + \dots \quad \text{............(D);}$$

then owing to the formula (10) and the following relation:

$$\frac{d^i}{dt^i}[e^{\phi(t)}] = m_i + m_{i+1}\frac{t}{1!} + m_{i+2}\frac{t^2}{2!} + \dots,$$

we shall have

$$m_i = N_0^{(i)},$$
$$m_i M_1 + m_{i+1} = N_1^{(i)},$$
$$m_i M_2 + 2M_1 m_{i+1} + m_{i+2} = N_2^{(i)},$$

or

$$\sum_0^\beta{}_j \binom{\beta}{j} M_{\beta-j} m_{i+j} = N_\beta^{(i)} \text{......... ....(E);}$$

hence $N_\beta^{(0)} = 0$, and if we put

$$\sum_0^a{}_i \binom{a}{i} M_{a-i} N_\beta^{(i)} = N_{a,\beta}^{(0)} \text{............................(14),}$$

the relation (13) can be reduced to the simpler form

$$\{\delta_1 s_a' \, \delta_1 s_b'\} = \frac{1}{N} N_{a,b}^{(0)} \quad \text{............................(F).}$$

Now, by multiplying the relation (D) by $\frac{u_i}{i!}$ and taking the sum from $i = 0$ to infinity, we obtain

$$e^{-\phi(t)} \sum_0^\infty{}_i \frac{u_i}{i!} \cdot \frac{d^i}{dt^i}[e^{\phi(t)}] = \sum_0^\infty{}_\beta \sum_0^\infty{}_i N_\beta^{(i)} \frac{t^\beta}{\beta!} \cdot \frac{u^i}{i!};$$

hence

$$e^{\phi(t+u) - \phi(t)} = \sum_0^\infty{}_\beta \sum_0^\infty{}_i N_\beta^{(i)} \frac{t^\beta}{\beta!} \cdot \frac{u^i}{i!} \quad \text{......................(G).}$$

That series gives by its coefficients the values of $N_\beta^{(i)}$. Multiplying the relation (G) by

$$e^{-\phi(u)} = 1 + M_1 \frac{u}{1!} + M_2 \frac{u^2}{2!} + \dots + M_p \frac{u^p}{p!} + \dots,$$

we obtain

$$e^{\phi(t+u) - \phi(t) - \phi(u)} = \sum_0^\infty{}_a \sum_0^\infty{}_b N_{a,b}^{(0)} \frac{t^a}{a!} \cdot \frac{u^b}{b!} \quad \text{....................(H),}$$

a relation which enables us to calculate the first approximation of any second order product moment $\{\delta_1 s_a' \, \delta_1 s_b'\}$ in terms of the semi-invariants of the parent population.

5. Let us now pass to the mean third order product; we shall have

$$\{\delta_1 s_a' \, \delta_1 s_b' \, \delta_1 s_c'\} = \left\{\left[\sum_1^a{}_i \binom{a}{i} M_{a-i}\,\delta m_i'\right]\left[\sum_1^b{}_j \binom{b}{j} M_{b-j}\,\delta m_j'\right]\left[\sum_1^c{}_k \binom{c}{k} M_{c-k}\,\delta m_k'\right]\right\}$$

$$= \sum_1^a{}_i \sum_1^b{}_j \sum_1^c{}_k \binom{a}{i}\binom{b}{j}\binom{c}{k} M_{a-i}\,M_{b-j}\,M_{c-k}\,\{\delta m_i'\,\delta m_j'\,\delta m_k'\},$$

and since

$$\{\delta m_i'\,\delta m_j'\,\delta m_k'\} = \frac{1}{N^2}(m_{i+j+k} - m_i m_{j+k} - m_j m_{k+i} - m_k m_{i+j} + 2m_i m_j m_k)$$

we shall obtain

$$\{\delta_1 s_a' \, \delta_1 s_b' \, \delta_1 s_c'\} = \frac{1}{N^2}\left[\sum_1^a{}_i \sum_1^b{}_j \sum_1^c{}_k \binom{a}{i}\binom{b}{j}\binom{c}{k} M_{a-i} M_{b-j} M_{c-k} m_{i+j+k}\right.$$

$$- \sum_1^a{}_i \binom{a}{i} M_{a-i} m_i \sum_1^b{}_j \sum_1^c{}_k \binom{b}{j}\binom{c}{k} M_{b-j} M_{c-k} m_{j+k}$$

$$- \sum_1^b{}_j \binom{b}{j} M_{b-j} m_j \sum_1^c{}_k \sum_1^a{}_i \binom{c}{k}\binom{a}{i} M_{c-k} M_{a-i} m_{k+i}$$

$$- \sum_1^c{}_k \binom{c}{k} M_{c-k} m_k \sum_1^a{}_i \sum_1^b{}_j \binom{a}{i}\binom{b}{j} M_{a-i} M_{b-j} m_{i+j}$$

$$\left. + 2\sum_1^a{}_i \binom{a}{i} M_{a-i} m_i \sum_1^b{}_j \binom{b}{j} M_{b-j} m_j \sum_1^c \binom{c}{k} M_{c-k} m_k\right] \quad \text{...(15).}$$

According to the relation (14) let us put

$$\sum_0^\beta{}_j \binom{\beta}{j} M_{\beta-j}\, N_\gamma^{(i+j)} = N_{\beta,\gamma}^{(i)} \quad \text{.......................(E}^{\text{bis}}\text{),}$$

and

$$\sum_0{}_j \binom{\beta}{j} M_{\beta-j}\, N_{\gamma,\delta}^{(i+j)} = N_{\beta,\gamma,\delta}^{(i)} \quad \text{.......................(E}^{\text{ter}}\text{).}$$

By using these relations and the previous ones, (C) and (E), the formula (15) becomes

$$\{\delta_1 s_a' \, \delta_1 s_b' \, \delta_1 s_c'\} = \frac{1}{N^2} N_{a,b,c}^{(0)} \quad \text{.......................(F}^{\text{bis}}\text{).}$$

But in order to have an approximation as far as the second power of $\frac{1}{N}$ we must calculate the fourth order product moment also, for

$$\{\delta m_i'\,\delta m_j'\,\delta m_k'\,\delta m_l'\}$$

$$= \frac{1}{N^3}(m_{i+j+k+l} - \mathbf{S} m_i m_{j+k+l} + \mathbf{S} m_{i+j} m_{k+l})$$

$$+ \frac{N-2}{N^3}(\mathbf{S} m_{i+j} m_{k+l} - \mathbf{S} m_i m_j m_{k+l} + 3 m_i m_j m_k m_l),$$

which gives

$$\{\delta_1 s_a' \, \delta_1 s_b' \, \delta_1 s_c' \, \delta_1 s_d'\} = \frac{1}{N^3}\left[\sum_1^a{}_i \sum_1^b{}_j \sum_1^c{}_k \binom{a}{i}\binom{b}{j}\binom{c}{k} M_{a-i} M_{b-j} M_{c-k} N_d^{(i+j+k)}\right.$$

$$+ M_a \sum_1^b{}_j \sum_1^c{}_k \binom{b}{j}\binom{c}{k} M_{b-j} M_{c-k} N_d^{(j+k)} + M_b \sum_1^c{}_k \sum_1^d{}_l \binom{c}{k}\binom{d}{l} M_{c-k} M_{d-l} N_a^{(k+l)}$$

$$+ M_c \sum_1^d{}_l \sum_1^a{}_i \binom{d}{l}\binom{a}{i} M_{d-l} M_{a-i} N_b^{(l+i)} + M_a M_c \sum_1^b{}_j \binom{b}{j} M_{b-j} N_d^{(j)}$$

$$+ M_c M_d \sum_{i=1}^{a} \binom{a}{i} M_{a-i} N_b^{(i)} + M_d M_b \sum_{k=1}^{c} \binom{c}{k} M_{c-k} N_a^{(k)}$$

$$+ \frac{N-1}{N^3}\left[\sum_{i=1}^{a} \binom{a}{i} M_{a-i} N_b^{(i)} \cdot \sum_{k=1}^{c} \binom{c}{k} M_{c-k} N_d^{(k)} + \sum_{j=1}^{b} \binom{b}{j} M_{b-j} N_c^{(j)} \cdot \sum_{l=1}^{d} \binom{d}{l} M_{d-l} N_a^{(l)}\right.$$

$$\left.+ \sum_{k=1}^{c} \binom{c}{k} M_{c-k} N_a^{(k)} \cdot \sum_{j=1}^{b} \binom{b}{j} M_{b-j} N_d^{(j)}\right] \quad \ldots\ldots\ldots\ldots(16).$$

In the same way as for the formulae (F) and (F^{bis}) by putting

$$\sum_{i=1}^{a} \binom{a}{i} M_{a-i} N^{(i)}_{\beta,\gamma,\delta} = N^{(0)}_{a,\beta,\gamma,\delta}$$

the relation (16) becomes

$$\{\delta_1 s_a' \, \delta_1 s_b' \, \delta_1 s_c' \, \delta_1 s_d'\}$$

$$= \frac{1}{N^3} N^{(0)}_{a,b,c,d} + \frac{N-1}{N^3}\left[N^{(0)}_{a,b} N^{(0)}_{c,d} + N^{(0)}_{a,c} N^{(0)}_{b,d} + N^{(0)}_{a,d} N^{(0)}_{b,c}\right] \ldots (F^{ter}).$$

Now let us differentiate the relation (G) i times with respect to u:

$$e^{-\phi(t)} \frac{\partial^i}{\partial u^i}\left[e^{\phi(t+u)}\right] = \sum_{\gamma=0}^{\infty} \frac{t^\gamma}{\gamma!} \sum_{j=0}^{\infty} N_\gamma^{(i+j)} \frac{u^j}{j!} \ldots\ldots\ldots\ldots(17),$$

hence, by considering the relation (E^{bis}), we deduce

$$e^{-\phi(t)-\phi(u)} \frac{\partial^i}{\partial u^i}\left[e^{\phi(t+u)}\right] = \sum_{\beta=0}^{\infty} \sum_{\gamma=0}^{\infty} N^{(i)}_{\beta,\gamma} \frac{u^\beta}{\beta!} \cdot \frac{t^\gamma}{\gamma!} \quad \ldots\ldots\ldots\ldots(D^{bis}),$$

and further

$$e^{\phi(t+u+v)-\phi(t)-\phi(u)} = \sum_{\beta=0}^{\infty} \sum_{\gamma=0}^{\infty} \sum_{i=0}^{\infty} N^{(i)}_{\beta,\gamma} \frac{u^\beta}{\beta!} \cdot \frac{t^\gamma}{\gamma!} \cdot \frac{v^i}{i!} \quad \ldots\ldots\ldots\ldots(G^{bis}).$$

If we multiply that relation by $e^{-\phi(v)}$ we finally obtain

$$e^{\phi(t+u+v)-\phi(t)-\phi(u)-\phi(v)} = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} N^{(0)}_{a,b,c} \frac{t^a}{a!} \cdot \frac{u^b}{b!} \cdot \frac{v^c}{c!} \ldots\ldots\ldots(H^{bis}).$$

This relation is to the third order product moment what (H) is to the second, i.e. the series on the right-hand side enables us to calculate the first approximation of $\{\delta_1 s_a' \, \delta_1 s_b' \, \delta_1 s_c'\}$.

The reader can now repeat the above calculation for the case of four semi-invariants. The result is the following formula:

$$e^{\phi(t+u+v+w)-\phi(t)-\phi(u)-\phi(v)-\phi(w)} = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \sum_{d=0}^{\infty} N^{(0)}_{a,b,c,d} \frac{t^a}{a!} \cdot \frac{u^b}{b!} \cdot \frac{v^c}{c!} \cdot \frac{w^d}{d!} \ldots (H^{ter}),$$

which gives the first term on the right-hand side of (F^{ter}), the expressions involved by the coefficient of $\frac{N-1}{N^3}$ being given by (H).

6. Now one may proceed and define by induction the general expression $N^{(i)}_{a_1,a_2,a_3,\ldots a_n}$, which is given by the coefficients of the series

$$e^{\phi(u_1+u_2+u_3+\ldots+u_{n+1})-\phi(u_1)-\phi(u_2)-\ldots-\phi(u_n)}$$

$$= \sum_{a_1=0}^{\infty} \sum_{a_2=0}^{\infty} \ldots \sum_{a_n=0}^{\infty} \sum_{i=0}^{\infty} N^{(i)}_{a_1,a_2,\ldots a_n} \frac{u_1^{a_1}}{a_1!} \cdot \frac{u_2^{a_2}}{a_2!} \ldots \frac{u_n^{a_n}}{a_n!} \cdot \frac{u_{n+1}^{i}}{i!} \quad \ldots\ldots\ldots(G^{ter}).$$

Still by induction we can write down the following table of formulae:

$$\sum_1^a {}_i \binom{a}{i} M_{a-i} m_{i+j} = N_a^{(j)} - M_a m_j,$$

$$\sum_1^a {}_i \sum_1^b {}_j \binom{a}{i}\binom{b}{j} M_{a-i} M_{b-j} m_{i+j+k}$$
$$= N_{a,b}^{(k)} - M_a N_b^{(k)} - M_b N_a^{(k)} + M_a M_b m_k,$$

$$\sum_1^a {}_i \sum_1^b {}_j \sum_1^c {}_k \binom{a}{i}\binom{b}{j}\binom{c}{k} M_{a-i} M_{b-j} M_{c-k} m_{i+j+k+l}$$
$$= N_{a,b,c}^{(l)} - M_a N_{b,c}^{(l)} - M_b N_{c,a}^{(l)} - M_c N_{a,b}^{(l)} + M_b M_c N_a^{(l)}$$
$$+ M_c M_a N_b^{(l)} + M_a M_b N_c^{(l)} - M_a M_b M_c m_l,$$

$$\sum_1^a {}_i \sum_1^b {}_j \sum_1^c {}_k \sum_1^d {}_l \binom{a}{i}\binom{b}{j}\binom{c}{k}\binom{d}{l} M_{a-i} M_{b-j} M_{c-k} M_{d-l} m_{i+j+k+l+p}$$
$$= N_{a,b,c,d}^{(p)} - \underset{(4)}{\mathbf{S}} M_a N_{b,c,d}^{(p)} + \underset{(6)}{\mathbf{S}} M_a M_b N_{c,d}^{(p)} - \underset{(4)}{\mathbf{S}} M_a M_b M_c N_d^{(p)}$$
$$+ M_a M_b M_c M_d m_p,$$

$$\sum_1^{a_1} {}_{i_1} \sum_1^{a_2} {}_{i_2} \cdots \sum_1^{a_n} {}_{i_n} \binom{a_1}{i_1}\binom{a_2}{i_2}\cdots\binom{a_n}{i_n} M_{a_1-i_1} M_{a_2-i_2} \cdots M_{a_n-i_n} m_{i_1+i_2+\ldots+i_n+k}$$
$$= N_{a_1,a_2\ldots a_n}^{(k)} - \underset{\binom{n}{1}}{\mathbf{S}} M_{a_1} N_{a_2,\ldots a_n}^{(k)} + \underset{\binom{n}{2}}{\mathbf{S}} M_{a_1} M_{a_2} N_{a_3,\ldots a_n}^{(k)} - \ldots$$
$$+ (-1)^j \underset{\binom{n}{j}}{\mathbf{S}} M_{a_1} M_{a_2} \ldots M_{a_j} N_{a_{j+1},\ldots a_n}^{(k)} + \ldots + (-1)^n M_{a_1} M_{a_2} \ldots M_{a_n} m_k \quad \ldots\ldots\ldots\text{(J)},$$

where $\underset{(q)}{\mathbf{S}}$ stands for a sum of q terms.

The formulae (G^{ter}) and (J) enable us to carry out the calculation of the first approximation of any product moment of the $\delta_1 s_p'$'s, provided that we have already found out the formula giving the corresponding product moment of the $\delta m_q'$'s.

By the method, which I shall give later (§ 8), we obtain

$$\{\delta m_i' \delta m_j' \delta m_k' \delta m_l' \delta m_p'\}$$
$$= \frac{1}{N^4}\Big[m_{i+j+k+l+p} - \underset{(5)}{\mathbf{S}} m_i m_{j+k+l+p} - \underset{(10)}{\mathbf{S}} m_{i+j} m_{k+l+p}$$
$$+ 2\underset{(10)}{\mathbf{S}} m_i m_j m_{k+l+p} + 2\underset{(15)}{\mathbf{S}} m_i m_{j+k} m_{l+p}$$
$$- 6\underset{(10)}{\mathbf{S}} m_i m_j m_k m_{l+p} + 24 m_i m_j m_k m_l m_p\Big]$$
$$+ \frac{1}{N^3}\Big[\underset{(10)}{\mathbf{S}} m_{i+j} m_{k+l+p} - \underset{(10)}{\mathbf{S}} m_i m_j m_{k+l+p} - 2\underset{(15)}{\mathbf{S}} m_i m_{j+k} m_{l+p}$$
$$+ 5\underset{(10)}{\mathbf{S}} m_i m_j m_k m_{l+p} - 20 m_i m_j m_k m_l m_p\Big] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(18)},$$

$$\{\delta m_i' \delta m_j' \delta m_k' \delta m_l' \delta m_p' \delta m_q'\}$$

$$= \frac{1}{N^5} \Big[m_{i+j+k+l+p+q} - \underset{(6)}{\mathbf{S}} m_i m_{j+k+l+p+q} - \underset{(15)}{\mathbf{S}} m_{i+j} m_{k+l+p+q}$$
$$- \underset{(10)}{\mathbf{S}} m_{i+j+k} m_{l+p+q} + 2 \underset{(15)}{\mathbf{S}} m_i m_j m_{k+l+p+q} + 2 \underset{(60)}{\mathbf{S}} m_i m_{j+k} m_{l+p+q}$$
$$+ 2 \underset{(15)}{\mathbf{S}} m_{i+j} m_{k+l} m_{p+q} - 6 \underset{(20)}{\mathbf{S}} m_i m_j m_k m_{l+p+q} - 6 \underset{(45)}{\mathbf{S}} m_i m_j m_{k+l} m_{p+q}$$
$$+ 24 \underset{(15)}{\mathbf{S}} m_i m_j m_k m_l m_{p+q} - 120 m_i m_j m_k m_l m_p m_q \Big]$$
$$+ \frac{1}{N^4} \Big[\underset{(15)}{\mathbf{S}} m_{i+j} m_{k+l+p+q} + \underset{(10)}{\mathbf{S}} m_{i+j+k} m_{l+p+q} - \underset{(15)}{\mathbf{S}} m_i m_j m_{k+l+p+q}$$
$$- 3 \underset{(15)}{\mathbf{S}} m_{i+j} m_{k+l} m_{p+q} - 2 \underset{(60)}{\mathbf{S}} m_i m_{j+k} m_{l+p+q} + 5 \underset{(20)}{\mathbf{S}} m_i m_j m_k m_{l+p+q}$$
$$+ 7 \underset{(45)}{\mathbf{S}} m_i m_j m_{k+l} m_{p+q} - 26 \underset{(15)}{\mathbf{S}} m_i m_j m_k m_l m_{p+q} + 130 m_i m_j m_k m_l m_p m_q \Big]$$
$$+ \frac{1}{N^3} \Big[\underset{(15)}{\mathbf{S}} m_{i+j} m_{k+l} m_{p+q} - \underset{(45)}{\mathbf{S}} m_i m_j m_{k+l} m_{p+q} + 3 \underset{(15)}{\mathbf{S}} m_i m_j m_k m_l m_{p+q}$$
$$- 15 m_i m_j m_k m_l m_p m_q \Big] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (19).$$

We have now everything which is necessary for proceeding to the calculation of the fifth and sixth order product moment of the $\delta_1 s_p''$s. We shall make one remark only concerning this calculation: one should not trouble too much about the sums $\mathbf{S}$, for the expressions involved by the formulae (J), (18) and (19) being symmetrical, the easiest way to deal with them is to count the number of terms which are analogous, independently of the number of different ways in which these may be obtained.

Let us take for an example the calculation of the coefficient of $\frac{1}{N^3}$ on the right-hand side of (18). We shall have

$$\underset{(10)}{\mathbf{S}}\, m_{i+j} m_{k+l+p} - \underset{(10)}{\mathbf{S}}\, m_i m_j m_{k+l+p}$$
$$\rightarrow \underset{(10)}{\mathbf{S}}\, N^{(0)}_{a,b} (N^{(0)}_{c,d,e} - \underset{(3)}{\mathbf{S}}\, M_c N^{(0)}_{d,e} - M_c M_d M_e)$$
$$= \underset{(10)}{\mathbf{S}}\, N^{(0)}_{a,b} N^{(0)}_{c,d,e} - \underset{(30)}{\mathbf{S}}\, M_a N^{(0)}_{b,c} N^{(0)}_{d,e} - \underset{(10)}{\mathbf{S}}\, M_a M_b M_c N^{(0)}_{d,e},$$

$$\underset{(15)}{\mathbf{S}}\, m_i m_{j+k} m_{l+p}$$
$$\rightarrow - \underset{(15)}{\mathbf{S}}\, M_a N^{(0)}_{b,c} N^{(0)}_{d,e} - \underset{(30)}{\mathbf{S}}\, M_a M_b M_c N^{(0)}_{d,e} - 15 M_a M_b M_c M_d M_e$$

$$\underset{(10)}{\mathbf{S}}\, m_i m_j m_k m_{l+p} \rightarrow - \underset{(10)}{\mathbf{S}}\, M_a M_b M_c N^{(0)}_{d,e} - 10 M_a M_b M_c M_d M_e;$$

hence the coefficient considered is $\underset{(10)}{\mathbf{S}}\, N^{(0)}_{a,b} N^{(0)}_{c,d,e}$.

In the same way we obtain

$$\{\delta_1 s_a' \delta_1 s_b' \delta_1 s_c' \delta_1 s_d' \delta_1 s_e'\} = \frac{1}{N^4} N^{(0)}_{a,b,c,d,e} + \frac{N-1}{N^4} \underset{(10)}{\mathbf{S}}\, N^{(0)}_{a,b} N^{(0)}_{c,d,e} \quad \ldots\ldots (\mathrm{F^{iv}})$$

and

$$\{\delta_1 s_a' \delta_1 s_b' \delta_1 s_c' \delta_1 s_d' \delta_1 s_e' \delta_1 s_f'\}$$

$$= \frac{1}{N^5} N_{a,b,c,d,e,f} + \frac{N-1}{N^5} \Big(\underset{(15)}{\mathbf{S}} N^{(0)}_{a,b} N^{(0)}_{c,d,e,f} + \underset{(10)}{\mathbf{S}} N^{(0)}_{a,b,c} N^{(0)}_{d,e,f} \Big)$$

$$+ \frac{(N-1)(N-2)}{N^5} \underset{(15)}{\mathbf{S}} N^{(0)}_{a,b} N^{(0)}_{c,d} N^{(0)}_{e,f} \quad \text{.........(F')}.$$

Let $L^{(1)}_{a,b\,c,\dots l}$ be the semi-invariant corresponding to the product moment

$$\{\delta_1 s_a' \delta_1 s_b' \delta_1 s_c' \dots \delta_1 s_l'\}.$$

We shall have

$$L^{(1)}_{a,b} = \frac{1}{N} N^{(0)}_{a,b},$$

$$L^{(1)}_{a,b,c} = \frac{1}{N^2} N^{(0)}_{a,b,c},$$

$$L^{(1)}_{a,b,c,d} = \frac{1}{N^3} \Big(N^{(0)}_{a,b,c,d} - \underset{(3)}{\mathbf{S}} N^{(0)}_{a,b} N^{(0)}_{c,d} \Big),$$

$$L^{(1)}_{a,b,c,d,e} = \frac{1}{N^4} \Big(N^{(0)}_{a,b,c,d,e} - \underset{(10)}{\mathbf{S}} N^{(0)}_{a,b} N^{(0)}_{c,d,e} \Big),$$

$$L^{(1)}_{a,b,c,d,e,f} = \frac{1}{N^5} \Big(N^{(0)}_{a,b,c,d,e,f} - \underset{(15)}{\mathbf{S}} N^{(0)}_{a,b} N^{(0)}_{c,d,e,f}$$

$$- \underset{(10)}{\mathbf{S}} N^{(0)}_{a,b,c} N^{(0)}_{d,e,f} + 2 \underset{(15)}{\mathbf{S}} N^{(0)}_{a,b} N^{(0)}_{c,d} N^{(0)}_{e,f} \Big) \quad \text{......(20)}.$$

The Associated Functions of Random Distributions.

7. The form of the right-hand sides of the formulae (20) led me to try to find the general expression of $L^{(1)}_{a,b,c,\dots l}$.

In order to work out this general problem I shall introduce new functions in connection with the distribution of random variables.

Supposing that we have a set of variables

$$y_0, y_1, y_2, \dots y_t \quad \text{.................................(21)}.$$

The following function,

$$\alpha_p(u_1, u_2, \dots u_p) = \sum_1^t {}_{i_1} \sum_1^t {}_{i_2} \dots \sum_1^t {}_{i_p} \{y_{i_1} y_{i_2} \dots y_{i_p}\} \frac{u_1^{i_1}}{i_1!} \cdot \frac{u_2^{i_2}}{i_2!} \dots \frac{u_p^{i_p}}{i_p!} \quad \text{......(22)},$$

will be called *the associated function with the pth order product moments of the set* (21). In the same way I define $\beta_p(u_1, u_2, u_3, \dots u_p)$, *the associated function with the pth order semi-invariants.*

Since we may easily establish the formulae

$$s_{111\dots1} = m_{111\dots1} - \mathbf{SS} m_{100\dots0} m_{011\dots1}$$

$$+ 2!\, \mathbf{SS} m_{100\dots0} m_{010\dots0} m_{0011\dots1}$$

$$- 3!\, \mathbf{SS} m_{10\dots0} m_{010\dots0} m_{0010\dots0} m_{00011\dots1}$$

$$+ \dots \quad \text{...(23)},$$

and $$m_{11\ldots 1} = s_{11\ldots 1} + \mathbf{SS} s_{100\ldots 0} s_{011\ldots 1}$$
$$+ \mathbf{SS} s_{10\ldots 0} s_{010\ldots 0} s_{001\ldots 1}$$
$$+ \mathbf{SS} s_{10\ldots 0} s_{010\ldots 0} s_{0010\ldots 0} s_{0001\ldots 1}$$
$$+ \ldots \ldots\ldots\ldots\ldots\ldots\ldots (24),$$

we shall have the following relations between the α's and β's:

$$\beta_p(u_1, u_2, \ldots u_p) = \alpha_p(u_1, u_2, \ldots u_p) - \mathbf{SS}\alpha_k(u_{i_1}, u_{i_2}, \ldots u_{i_k})\,\alpha_{p-k}(u_{i_{k+1}}, \ldots u_{i_p})$$
$$+ 2!\,\mathbf{SS}\alpha_k(u_{i_1}, u_{i_2}, \ldots u_{i_k})\,\alpha_l(u_{i_{k+1}}, \ldots u_{i_{k+l}})\,\alpha_{p-k-l}(u_{i_{k+l+1}}, u_{i_{k+l+2}}, \ldots u_{i_p})$$
$$- 3!\,\mathbf{SS}\alpha_k(u_{i_1}, u_{i_2}, \ldots u_{i_k})\,\alpha_l(u_{i_{k+1}}, \ldots u_{i_{k+l}})\,\alpha_m(u_{i_{k+l+1}}, \ldots u_{i_{k+l+m}})$$
$$\alpha_{p-k-l-m}(u_{i_{k+l+m+1}}, \ldots u_{i_p})$$
$$+ \ldots \ldots\ldots\ldots\ldots\ldots\ldots (25),$$

$$\alpha_p(u_1, u_2, \ldots u_p) = \beta_p(u_1, u_2, \ldots u_p) + \mathbf{SS}\beta_k(u_{i_1}, u_{i_2}, \ldots u_{i_k})\,\beta_{p-k}(u_{i_{k+1}}, \ldots u_{i_p})$$
$$+ \mathbf{SS}\beta_k(u_{i_1}, u_{i_2}, \ldots u_{i_k})\,\beta_l(u_{i_{k+1}}, \ldots u_{i_{k+l}})\,\beta_{p-k-l}(u_{i_{k+l+1}}, \ldots u_{i_p})$$
$$+ \mathbf{SS}\beta_k(u_{i_1}, u_{i_2}, \ldots u_{i_k})\,\beta_l(u_{i_{k+1}}, \ldots u_{i_{k+l}})\,\beta_m(u_{i_{k+l+1}}, \ldots u_{i_{k+l+m}})$$
$$\beta_{p-k-l-m}(u_{i_{k+l+m+1}}, \ldots u_{i_p})$$
$$+ \ldots \ldots\ldots\ldots\ldots\ldots\ldots (26).$$

I think that the associated functions will allow us to treat many questions with regard to an infinite series of random variables—like the sampling moments or semi-invariants—in a more effective manner than the corresponding characteristic function does. Moreover, I have already dealt with such functions in the previous chapter, and their usefulness causes me to make special mention of them.

Let $$g(t) = y_0 + y_1\frac{t}{1!} + y_2\frac{t^2}{2!} + \ldots + y_p\frac{t^p}{p!} + \ldots$$

be the generant function of the y's, and $F dv$ the elementary probability of the event $(y_0, y_1, y_2, \ldots y_p, \ldots)$.

We shall have

$$\alpha_1(t_1) = \int_{(D)} F g(t_1)\,dv,$$
$$\alpha_2(t_1, t_2) = \int_{(D)} F g(t_1)\,g(t_2)\,dv,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\alpha_p(t_1, t_2, \ldots t_p) = \int_{(D)} F g(t_1)\,g(t_2)\ldots g(t_p)\,dv,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots (27);$$

hence it follows

$$\psi(t_1, t_2, \ldots t_n, \ldots) = \int_{(D)} F e^{\lambda[g(t_1)+g(t_2)+\cdots+g(t_n)+\cdots]}\,dv$$
$$= 1 + \frac{\lambda}{1!}\Sigma_i \alpha_1(t_i) + \frac{\lambda^2}{2!}\Sigma_{i,j}\alpha_2(t_i, t_j) + \ldots$$
$$= e^{\frac{\lambda}{1!}\Sigma_i\beta_1(t_i) + \frac{\lambda^2}{2!}\Sigma_{i,j}\beta_2(t_i, t_j) + \ldots} \ldots\ldots\ldots\ldots (28).$$

A similar relation may be established for a multivariate distribution. If

$$g(t, u) = y_{0,0} + y_{1,0}\frac{t}{1!} + y_{0,1}\frac{u}{1!} + y_{2,0}\frac{t^2}{2!} + y_{1,1}\frac{t}{1!}\cdot\frac{u}{1!} + y_{0,2}\frac{u^2}{2!} + \dots$$

is the generant function of $y_{i,j}$ we shall have

$$\psi\begin{pmatrix} t_1, t_2, \dots t_p, \dots \\ u_1, u_2, \dots u_p, \dots \end{pmatrix} = \int_{(D)} F e^{\lambda[g(t_1, u_1) + g(t_2, u_2) + \dots + g(t_n, u_n) + \dots]}\, dv$$

$$= 1 + \frac{\lambda}{1!}\Sigma_i \alpha_1\begin{pmatrix} t_i \\ u_i \end{pmatrix} + \frac{\lambda^2}{2!}\Sigma_{i,j}\alpha_2\begin{pmatrix} t_i, t_j \\ u_i, u_j \end{pmatrix} + \frac{\lambda^3}{3!}\Sigma_{i,j,k}\alpha_3\begin{pmatrix} t_i, t_j, t_k \\ u_i, u_j, u_k \end{pmatrix} + \dots$$

$$= e^{\frac{\lambda}{1!}\Sigma_i\beta_1\binom{t_i}{u_i} + \frac{\lambda^2}{2!}\Sigma_{i,j}\beta_2\binom{t_i, t_j}{u_i, u_j} + \dots} \quad \dots\dots\dots(28^{\text{bis}});$$

hence it follows

$$\beta_p\begin{pmatrix} t_1, t_2, \dots t_p \\ u_1, u_2, \dots u_p \end{pmatrix} = \alpha_p\begin{pmatrix} t_1, t_2, \dots t_p \\ u_1, u_2, \dots u_p \end{pmatrix} - \mathbf{SS}\alpha_k\begin{pmatrix} t_{i_1}, t_{i_2}, \dots t_{i_k} \\ u_{i_1}, u_{i_2}, \dots u_{i_k} \end{pmatrix}\alpha_{p-k}\begin{pmatrix} t_{i_{k+1}}, \dots t_{i_p} \\ u_{i_{k+1}}, \dots u_{i_p} \end{pmatrix}$$

$$+ 2!\,\mathbf{SS}\alpha_k\begin{pmatrix} t_{i_1}, \dots t_{i_k} \\ u_{i_1}, \dots u_{i_k} \end{pmatrix}\alpha_l\begin{pmatrix} t_{i_{k+1}}, \dots t_{i_{k+l}} \\ u_{i_{k+1}}, \dots u_{i_{k+l}} \end{pmatrix}\alpha_{p-k-l}\begin{pmatrix} t_{i_{k+l+1}}, \dots t_{i_p} \\ u_{i_{k+l+1}}, \dots u_{i_p} \end{pmatrix}$$

$$- \dots\dots\dots$$

and

$$\alpha_p\begin{pmatrix} t_1, t_2, \dots t_p \\ u_1, u_2, \dots u_p \end{pmatrix} = \beta_p\begin{pmatrix} t_1, t_2, \dots t_p \\ u_1, u_2, \dots u_p \end{pmatrix} + \mathbf{SS}\beta_k\begin{pmatrix} t_{i_1}, t_{i_2}, \dots t_{i_k} \\ u_{i_1}, u_{i_2}, \dots u_{i_k} \end{pmatrix}\beta_{p-k}\begin{pmatrix} t_{i_{k+1}}, \dots t_{i_p} \\ u_{i_{k+1}}, \dots u_{i_p} \end{pmatrix}$$

$$+ \mathbf{SS}\beta_k\begin{pmatrix} t_{i_1}, t_{i_2}, \dots t_{i_k} \\ u_{i_1}, u_{i_2}, \dots u_{i_k} \end{pmatrix}\beta_l\begin{pmatrix} t_{i_{k+1}}, \dots t_{i_{k+l}} \\ u_{i_{k+1}}, \dots u_{i_{k+l}} \end{pmatrix}\beta_{p-k-l}\begin{pmatrix} t_{i_{k+l+1}}, \dots t_{i_p} \\ u_{i_{k+l+1}}, \dots u_{i_p} \end{pmatrix}$$

$$+ \dots\dots\dots$$

These relations—which may be formal only—represent the generalisation of the characteristic function, and enable us to find out any associated function.

Associated Functions of the Distribution of Moments and Product Moments about a Fixed Origin.

8. I shall first make an application of the formula (28) to the correlated distribution of sampling moments about a fixed point*.

In that case $\quad g(t) = \frac{t}{1!}\delta m_1' + \frac{t^2}{2!}\delta m_2' + \dots + \frac{t^p}{p!}\delta m_p' + \dots$

$$= \frac{1}{N}\sum_1^N{}_i\,[e^{x_i t} - e^{\phi(t)}] \quad \dots\dots\dots(29),$$

where $e^{\phi(t)}$ is the characteristic function of the distribution of x. Thus

$$\psi(t_1, t_2, \dots t_n, \dots)$$

$$= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \dots \int_{-\infty}^{+\infty} f(x_1) f(x_2) \dots f(x_N)\, e^{\frac{\lambda}{N}\sum_1^\infty{}_r \sum_1^N{}_i [e^{x_i t_r} - e^{\phi(t_r)}]}\, dx_N$$

$$= [\psi(t_1, t_2, \dots t_n, \dots)]^N,$$

* A method for calculating the semi-invariants of the distribution of one moment only has been indicated by C. C. Craig, *loc. cit.* p. 13.

where

$$\psi(t_1, t_2, \ldots t_n, \ldots) = \int_{-\infty}^{+\infty} f(x)\, e^{\frac{\lambda}{N}\Sigma_r [e^{xt_r} - e^{\phi(t_r)}]}\, dx$$

$$= 1 + \frac{1}{2!}\cdot\frac{\lambda^2}{N^2}\Sigma_{i,j} ⟪e^{\phi(t_i)} - e^{\phi(t_i)}⟫ ⟪e^{\phi(t_j)} - e^{\phi(t_j)}⟫$$

$$+ \frac{1}{3!}\cdot\frac{\lambda^3}{N^3}\Sigma_{i,j,k} ⟪e^{\phi(t_i)} - e^{\phi(t_i)}⟫ ⟪e^{\phi(t_j)} - e^{\phi(t_j)}⟫ ⟪e^{\phi(t_k)} - e^{\phi(t_k)}⟫,$$

the double brackets standing for a symbolical multiplication in order to have

$$\mathrm{a}_2(t_1, t_2) = \frac{1}{N^2}[e^{\phi(t_1+t_2)} - e^{\phi(t_1)+\phi(t_2)}],$$

$$\mathrm{a}_3(t_1, t_2, t_3) = \frac{1}{N^3}[e^{\phi(t_1+t_2+t_3)} - e^{\phi(t_1)+\phi(t_2+t_3)} - e^{\phi(t_2)+\phi(t_3+t_1)} - e^{\phi(t_3)+\phi(t_1+t_2)} + 2e^{\phi(t_1)+\phi(t_2)+\phi(t_3)}],$$

$$\mathrm{a}_p(t_1, t_2, \ldots t_p) = \frac{1}{N^p}[e^{\phi(t_1+t_2+\cdots+t_p)} - \mathbf{S}\, e^{\phi(t_1)+\phi(t_2+t_3+\cdots+t_p)} + \mathbf{S}\, e^{\phi(t_1)+\phi(t_2)+\phi(t_3+t_4+\cdots+t_p)}$$

$$+ (-1)^{p-1}(p-1)\, e^{\phi(t_1)+\phi(t_2)+\cdots+\phi(t_p)}] \quad \ldots\ldots\ldots(30);$$

hence the β's connected with $\psi(t_1, t_2, t_3, \ldots t_n, \ldots)$ and giving the semi-invariants of the $\delta m_p'$'s will be

$$\mathrm{b}_2(t_1, t_2) = \frac{1}{N}[e^{\phi(t_1+t_2)} - e^{\phi(t_1)+\phi(t_2)}],$$

$$\mathrm{b}_3(t_1, t_2, t_3) = \frac{1}{N^2}[e^{\phi(t_1+t_2+t_3)} - \mathbf{S}\, e^{\phi(t_1)+\phi(t_2+t_3)} + 2e^{\phi(t_1)+\phi(t_2)+\phi(t_3)}],$$

$$\mathrm{b}_4(t_1, t_2, t_3, t_4) = \frac{1}{N^3}[e^{\phi(t_1+t_2+t_3+t_4)} - \mathbf{S}\, e^{\phi(t_1)+\phi(t_2+t_3+t_4)} - \mathbf{S}\, e^{\phi(t_1+t_2)+\phi(t_3+t_4)} + 2!\,\mathbf{S}\, e^{\phi(t_1)+\phi(t_2)+\phi(t_3+t_4)} - 3!\, e^{\phi(t_1)+\phi(t_2)+\phi(t_3)+\phi(t_4)}],$$

....

$$\mathrm{b}_p(t_1, t_2, \ldots t_p) = \frac{1}{N^{p-1}}[e^{\phi(t_1+t_2+\cdots+t_p)} - \mathbf{S}\mathbf{S}\, e^{\phi(t_1+t_2+\cdots+t_i)+\phi(t_{i+1}+\cdots+t_p)} + 2!\,\mathbf{S}\mathbf{S}\, e^{\phi(t_1+t_2+\cdots+t_i)+\phi(t_{i+1}+\cdots+t_{i+k})+\phi(t_{i+k+1}+\cdots+t_p)}$$

$$+ (-1)^{p-1}(p-1)!\, e^{\phi(t_1)+\phi(t_2)+\cdots+\phi(t_p)}],$$

....

$$\ldots\ldots\ldots(31).$$

The associated functions with the moments of the $\delta m_p'$'s are easily calculated either from (30) or from (31):

$$a_2(t_1, t_2) = \frac{1}{N}[e^{\phi(t_1+t_2)} - e^{\phi(t_1)+\phi(t_2)}],$$

$$a_3(t_1, t_2, t_3) = \frac{1}{N^2}[e^{\phi(t_1+t_2+t_3)} - \mathbf{S}\, e^{\phi(t_1)+\phi(t_2+t_3)} + 2e^{\phi(t_1)+\phi(t_2)+\phi(t_3)}],$$

$$a_4(t_1, t_2, t_3, t_4) = \frac{1}{N^3}\left[e^{\phi(t_1+t_2+t_3+t_4)} - \mathbf{S}\, e^{\phi(t_1)+\phi(t_2+t_3+t_4)}\right.$$
$$\left. + \mathbf{S}\, e^{\phi(t_1+t_2)+\phi(t_3+t_4)} - \mathbf{S}\, e^{\phi(t_1)+\phi(t_2)+\phi(t_3+t_4)} + 3e^{\phi(t_1)+\phi(t_2)+\phi(t_3)+\phi(t_4)}\right]$$
$$+ \frac{N-1}{N^3}\mathbf{S}\left[e^{\phi(t_1+t_2)} - e^{\phi(t_1)+\phi(t_2)}\right]\left[e^{\phi(t_3+t_4)} - e^{\phi(t_3)+\phi(t_4)}\right],$$

....

$$a_p(t_1, t_2, \dots t_p) = N\mathfrak{a}_p(t_1, t_2, \dots t_p)$$
$$+ N(N-1)\,\mathbf{S}\,\mathfrak{a}_k(t_1, t_2, \dots t_k)\,\mathfrak{a}_{p-k}(t_{k+1}, t_{k+2}, \dots t_p)$$
$$+ N(N-1)(N-2)\,\mathbf{S}\,\mathfrak{a}_k(t_1, t_2, \dots t_k)\,\mathfrak{a}_l(t_{k+1}, \dots t_{k+l})\,\mathfrak{a}_{p-l-k}(t_{k+l+1}, \dots t_p)$$
$$+ \dots \qquad \text{..............................} \qquad (32);$$

hence we may obtain the product moments of the $\delta m_p'$'s. (See above, formulae (18) and (19).) Their calculations present no difficulties, for the coefficient of $\frac{u_1^i}{i!}\cdot\frac{u_2^j}{j!}\cdot\frac{u_3^k}{k!}\dots$ in $e^{\phi(u_1+u_2+\dots)}$ is $m_{i+j+k+\dots}$.

9. The method which has been developed for finding the product moments and the semi-invariants of the distribution of sampling moments about a fixed origin is readily extended to the case of sampling product moments about a fixed origin of a multivariate distribution. In addition to being very useful, this problem is also a further illustration of the method of associated functions. I shall treat it for the case of a frequency function of two variates, after which the calculations for a frequency function of three or more variates will be straightforward.

We shall have

$$g(t, u) = \frac{t}{1!}\delta m'_{10} + \frac{u}{1!}\delta m'_{01} + \frac{t^2}{2!}\delta m'_{20} + \frac{t}{1!}\cdot\frac{u}{1!}\delta m'_{11} + \frac{u^2}{2!}\delta m'_{02}$$
$$+ \frac{t^3}{3!}\delta m'_{30} + \frac{t^2}{2!}\cdot\frac{u}{1!}\delta m'_{21} + \frac{t}{1!}\cdot\frac{u^2}{2!}\delta m'_{12} + \frac{u^3}{3!}\delta m'_{03} + \dots$$
$$= \frac{1}{N}\sum_1^N{}_i\left[e^{x_i t + y_i u} - e^{\phi\binom{t}{u}}\right];$$

hence
$$\psi\begin{pmatrix} t_1, & t_2, & \dots & t_p, & \dots \\ u_1, & u_2, & \dots & u_p, & \dots\end{pmatrix} = \left[\psi\begin{pmatrix} t_1, & t_2, & \dots & t_p, & \dots \\ u_1, & u_2, & \dots & u_p, & \dots\end{pmatrix}\right]^N,$$
where
$$\begin{pmatrix} t_1, & t_2, & \dots & t_p, & \dots \\ u_1, & u_2, & \dots & u_p, & \dots\end{pmatrix} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\, e^{\frac{\lambda}{N}\Sigma_r\left[e^{xt_r+yu_r} - e^{\phi\binom{t_r}{u_r}}\right]}dx\,dy$$
$$= 1 + \frac{\lambda^2}{2!}\Sigma_{i,j}\,\mathfrak{a}_2\begin{pmatrix} t_i, & t_j \\ u_i, & u_j\end{pmatrix} + \frac{\lambda^3}{3!}\Sigma_{i,j,k}\,\mathfrak{a}_3\begin{pmatrix} t_i, & t_j, & t_k \\ u_i, & u_j, & u_k\end{pmatrix} + \dots.$$

Developing the right-hand side of this relation we obtain the following formulae:

$$\mathfrak{a}_2\begin{pmatrix} t_1, & t_2 \\ u_1, & u_2\end{pmatrix} = \frac{1}{N^2}\left[e^{\phi\binom{t_1+t_2}{u_1+u_2}} - e^{\phi\binom{t_1}{u_1}+\phi\binom{t_2}{u_2}}\right],$$

$$\mathfrak{a}_3\begin{pmatrix} t_1, & t_2, & t_3 \\ u_1, & u_2, & u_3\end{pmatrix} = \frac{1}{N^3}\left[e^{\phi\binom{t_1+t_2+t_3}{u_1+u_2+u_3}} - \mathbf{S}\,e^{\phi\binom{t_1}{u_1}+\phi\binom{t_2+t_3}{u_2+u_3}} + 2e^{\phi\binom{t_1}{u_1}+\phi\binom{t_2}{u_2}+\phi\binom{t_3}{u_3}}\right],$$

$$a_p\begin{pmatrix}t_1, & t_2, & \dots & t_p\\ u_1, & u_2, & \dots & u_p\end{pmatrix} = \frac{1}{N^p}\Bigg[e^{\phi\begin{pmatrix}t_1+t_2+\dots+t_p\\ u_1+u_2+\dots+u_p\end{pmatrix}} - \mathbf{S}\, e^{\phi\begin{pmatrix}t_1\\ u_1\end{pmatrix}+\phi\begin{pmatrix}t_2+t_3+\dots+t_p\\ u_2+u_3+\dots+u^p\end{pmatrix}}$$
$$+\mathbf{S}\, e^{\phi\begin{pmatrix}t_1\\ u_1\end{pmatrix}+\phi\begin{pmatrix}t_2\\ u_2\end{pmatrix}+\phi\begin{pmatrix}t_3+t_4+\dots+t_p\\ u_3+u_4+\dots+u_p\end{pmatrix}}$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$+(-1)^{p-1}(p-1)\, e^{\phi\begin{pmatrix}t_1\\ u_1\end{pmatrix}+\phi\begin{pmatrix}t_2\\ u_2\end{pmatrix}+\phi\begin{pmatrix}t_3\\ u_3\end{pmatrix}+\dots+\phi\begin{pmatrix}t_p\\ u_p\end{pmatrix}}\Bigg]$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\dots\dots(30^{\text{bis}});$$

hence the α's and β's connected with $\psi\begin{pmatrix}t_1, & t_2, & \dots & t_n, & \dots\\ u_1, & u_2, & \dots & u_n, & \dots\end{pmatrix}$ will be

$$b_p\begin{pmatrix}t_1, & t_2, & \dots & t_p\\ u_1, & u_2, & \dots & u_p\end{pmatrix}$$
$$= \frac{1}{N^{p-1}}\Bigg[e^{\phi\begin{pmatrix}t_1+t_2+\dots+t_p\\ u_1+u_2+\dots+u_p\end{pmatrix}} - \mathbf{SS}\, e^{\phi\begin{pmatrix}t_1+t_2+\dots+t_k\\ u_1+u_2+\dots+u_k\end{pmatrix}+\phi\begin{pmatrix}t_{k+1}+\dots+t_p\\ u_{k+1}+\dots+u_p\end{pmatrix}}$$
$$+2!\,\mathbf{SS}\, e^{\phi\begin{pmatrix}t_1+t_2+\dots+t_k\\ u_1+u_2+\dots+u_k\end{pmatrix}+\phi\begin{pmatrix}t_{k+1}+\dots+t_{k+l}\\ u_{k+1}+\dots+u_{k+l}\end{pmatrix}+\phi\begin{pmatrix}t_{k+l+1}+\dots+t_p\\ u_{k+l+1}+\dots+u_p\end{pmatrix}}$$
$$+(-1)^{p-1}(p-1)!\, e^{\phi\begin{pmatrix}t_1\\ u_1\end{pmatrix}+\phi\begin{pmatrix}t_2\\ u_2\end{pmatrix}+\dots+\phi\begin{pmatrix}t_p\\ u_p\end{pmatrix}}\Bigg] \dots\dots\dots\dots(31^{\text{bis}}),$$

and

$$a_p\begin{pmatrix}t_1, & t_2, & \dots & t_p\\ u_1, & u_2, & \dots & u_p\end{pmatrix}$$
$$= N\mathrm{a}_p\begin{pmatrix}t_1, & t_2, & \dots & t_p\\ u_1, & u_2, & \dots & u_p\end{pmatrix} + N(N-1)\,\mathbf{SS}\,\mathrm{a}_k\begin{pmatrix}t_1, & t_2, & \dots & t_k\\ u_1, & u_2, & \dots & u_k\end{pmatrix}\mathrm{a}_{p-k}\begin{pmatrix}t_{k+1}, & \dots & t_p\\ u_{k+1}, & \dots & u_p\end{pmatrix}$$
$$+N(N-1)(N-2)\,\mathbf{SS}\,\mathrm{a}_k\begin{pmatrix}t_1, & t_2, & \dots & t_k\\ u_1, & u_2, & \dots & u_k\end{pmatrix}\mathrm{a}_l\begin{pmatrix}t_{k+1}, & \dots & t_{k+l}\\ u_{k+1}, & \dots & u_{k+l}\end{pmatrix}\mathrm{a}_{p-k-l}\begin{pmatrix}t_{k+l+1}, & \dots & t_p\\ u_{k+l+1}, & \dots & u_p\end{pmatrix}$$
$$+\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\dots\dots(32^{\text{bis}});$$

hence it follows at once:

$$\{\delta m'_{i_1,j_1}\delta m'_{i_2,j_2}\} = \frac{1}{N}(m_{i_1+i_2,j_1+j_2} - m_{i_1,j_1}m_{i_2,j_2}),$$
$$\{\delta m'_{i_1,j_1}\delta m'_{i_2,j_2}\delta m'_{i_3,j_3}\} = \frac{1}{N^2}(m_{i_1+i_2+i_3,j_1+j_2+j_3} - m_{i_1,j_1}m_{i_2+i_3,j_2+j_3}$$
$$- m_{i_2,j_2}m_{i_3+i_1,j_3+j_1} - m_{i_3,j_3}m_{i_1+i_2,j_1+j_2} + 2m_{i_1,j_1}m_{i_2,j_2}m_{i_3,j_3}),$$
$$\{\delta m'_{i_1,j_1}\delta m'_{i_2,j_2}\delta m'_{i_3,j_3}\delta m'_{i_4,j_4}\} = \frac{1}{N^3}(m_{i_1+i_2+i_3+i_4,j_1+j_2+j_3+j_4}$$
$$- \mathbf{S}\, m_{i_1,j_1}m_{i_2+i_3+i_4,j_2+j_3+j_4} + \mathbf{S}\, m_{i_1,j_1}m_{i_2,j_2}m_{i_3+i_4,j_3+j_4}$$
$$- 3m_{i_1,j_1}m_{i_2,j_2}m_{i_3,j_3}m_{i_4,j_4})$$
$$+ \frac{N-1}{N^3}\,\mathbf{S}\,(m_{i_1+i_2,j_1+j_2} - m_{i_1,j_1}m_{i_2,j_2})(m_{i_3+i_4,j_3+j_4} - m_{i_3,j_3}m_{i_4,j_4}),$$
$$\dots\dots(33).$$

Finally I shall draw the reader's attention to the fact that the formulae established in this section present a similarity in their forms. This property may have been noticed before, but it is through the method of associated functions that we have obtained an explanation of the fact. We observe then, that—owing to the similarity of these formulae—knowing one of them, we are certain to obtain those which are analogous to it by making the necessary modifications.

Further Results for the Distribution of the Semi-invariants.

10. Having obtained the general formulae for the distribution of sampling moments we are able to proceed to find out the general expression of $L^{(1)}_{a, b, c, \dots l}$, or, which is the same, the corresponding associated function.

If we put $$G_1(t) = \frac{t}{1!}\delta_1 s_1' + \frac{t^2}{2!}\delta_1 s_2' + \dots + \frac{t^p}{p!}\delta_1 s_p' + \dots,$$
according to the relations (9) and (29) we shall obtain
$$G_1(t) = e^{-\phi(t)} g(t) \qquad \dots\dots(34).$$

Let us denote by $A_p^{(1)}(t_1, t_2, t_3, \dots t_p)$ and $B_p^{(1)}(t_1, t_2, t_3, \dots t_p)$ respectively the associated function with the pth order product moment and pth order semi-invariant of the $\delta_1 s_p'$'s.

Owing to the relation (27) we shall have
$$A_p^{(1)}(t_1, t_2, \dots t_p) = e^{-\phi(t_1) - \phi(t_2) - \dots - \phi(t_p)} a_p(t_1, t_2, \dots t_p) \qquad \dots\dots(35),$$
where a_p is the associated function with the pth order product moment of the $\delta m_p'$'s.

Now by using the relations (25) we obtain further
$$B_p^{(1)}(t_1, t_2, \dots t_p) = e^{-\phi(t_1) - \phi(t_2) - \dots - \phi(t_p)} b_p(t_1, t_2, \dots t_p) \qquad \dots\dots(36),$$
where $b_p(t_1, t_2, \dots t_p)$ is given by (31).

The formulae (35) and (36) solve the problem which I proposed working out in the second section. The method followed in attaining this aim constitutes an example of the usefulness of the associated functions.

Let us put
$$\begin{aligned}\mathcal{A}_p(t_1, t_2, \dots t_p) &= N^p a_p(t_1, t_2, \dots t_p)\, e^{-\phi(t_1) - \phi(t_2) - \dots - \phi(t_p)} \\ &= e^{\phi(t_1+t_2+\dots+t_p) - \phi(t_1) - \phi(t_2) - \dots - \phi(t_p)} \\ &\quad - \mathbf{S}\, e^{\phi(t_2+t_3+\dots+t_p) - \phi(t_2) - \phi(t_3) - \dots - \phi(t_p)} \\ &\quad + \mathbf{S}\, e^{\phi(t_3+t_4+\dots+t_p) - \phi(t_3) - \phi(t_4) - \dots - \phi(t_p)} \\ &\quad + (-1)^{p-1}(p-1) \qquad \dots\dots(37).\end{aligned}$$

The relations (35) and (36) will give the following table of formulae:

$$A_2^{(1)}(t_1, t_2) = \frac{1}{N}\mathcal{A}_2(t_1, t_2),$$

$$A_3^{(1)}(t_1, t_2, t_3) = \frac{1}{N^2}\mathcal{A}_3(t_1, t_2, t_3),$$

$$A_4^{(1)}(t_1, t_2, t_3, t_4) = \frac{1}{N^3} \mathscr{A}_4(t_1, t_2, t_3, t_4) + \frac{N-1}{N^3} \underset{(3)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_2(t_3, t_4),$$

$$A_5^{(1)}(t_1, t_2, t_3, t_4, t_5) = \frac{1}{N^4} \mathscr{A}_5(t_1, t_2, t_3, t_4, t_5)$$
$$+ \frac{N-1}{N^4} \underset{(10)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_3(t_3, t_4, t_5),$$

$$A_6^{(1)}(t_1, t_2, t_3, t_4, t_5, t_6) = \frac{1}{N^5} \mathscr{A}_6(t_1, t_2, t_3, t_4, t_5, t_6)$$
$$+ \frac{N-1}{N^5} [\underset{(15)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_4(t_3, t_4, t_5, t_6) + \underset{(10)}{\mathbf{S}} \mathscr{A}_3(t_1, t_2, t_3) \mathscr{A}_3(t_4, t_5, t_6)]$$
$$+ \frac{(N-1)(N-2)}{N^5} \underset{(15)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_2(t_3, t_4) \mathscr{A}_2(t_5, t_6),$$

$$A_7(t_1, t_2, \ldots t\ , t_7) = \frac{1}{N^6} \mathscr{A}_7(t_1, t_2, \ldots t_6, t_7)$$
$$+ \frac{N-1}{N^6} [\underset{(21)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_5(t_3, t_4, \ldots t_7) + \underset{(35)}{\mathbf{S}} \mathscr{A}_3(t_1, t_2, t_3) \mathscr{A}_4(t_4, t_5, t_6, t_7)]$$
$$+ \frac{(N-1)(N-2)}{N^6} \underset{(105)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_2(t_3, t_4) \mathscr{A}_3(t_5, t_6, t_7),$$

$$A_8(t_1, t_2, \ldots t_7, t_8) = \frac{1}{N^7} \mathscr{A}_8(t_1, t_2, \ldots t_7, t_8)$$
$$+ \frac{N-1}{N^7} [\underset{(28)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_6(t_3, t_4, \ldots t_8) + \underset{(56)}{\mathbf{S}} \mathscr{A}_3(t_1, t_2, t_3) \mathscr{A}_5(t_4, t_5, \ldots t_8)$$
$$+ \underset{(35)}{\mathbf{S}} \mathscr{A}_4(t_1, t_2, t_3, t_4) \mathscr{A}_4(t_5, t_6, t_7, t_8)]$$
$$+ \frac{(N-1)(N-2)}{N^7} [\underset{(210)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_2(t_3, t_4) \mathscr{A}_4(t_5, t_6, t_7, t_8)$$
$$+ \underset{(280)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_3(t_3, t_4, t_5) \mathscr{A}_3(t_6, t_7, t_8)]$$
$$+ \frac{(N-1)(N-2)(N-3)}{N^7} \underset{(105)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_2(t_3, t_4) \mathscr{A}_2(t_5, t_6) \mathscr{A}_2(t_7, t_8) \quad \text{.(38)},$$

and

$$B_2^{(1)}(t_1, t_2) = \frac{1}{N} \mathscr{A}_2(t_1, t_2),$$

$$B_3^{(1)}(t_1, t_2, t_3) = \frac{1}{N^2} \mathscr{A}_3(t_1, t_2, t_3),$$

$$B_4^{(1)}(t_1, t_2, t_3, t_4) = \frac{1}{N^3} [\mathscr{A}_4(t_1, t_2, t_3, t_4) - \underset{(3)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_2(t_3, t_4)],$$

$$B_5^{(1)}(t_1, t_2, t_3, t_4, t_5) = \frac{1}{N^4} [\mathscr{A}_5(t_1, t_2, \ldots t_5) - \underset{(10)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_3(t_3, t_4, t_5)],$$

$$B_6^{(1)}(t_1, t_2, t_3, t_4, t_5, t_6) = \frac{1}{N^5} [\mathscr{A}_6(t_1, t_2, \ldots t_6) - \underset{(15)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_4(t_3, t_4, t_5, t_6)$$
$$- \underset{(10)}{\mathbf{S}} \mathscr{A}_3(t_1, t_2, t_3) \mathscr{A}_3(t_4, t_5, t_6) + 2!\ \underset{(15)}{\mathbf{S}} \mathscr{A}_2(t_1, t_2) \mathscr{A}_2(t_3, t_4) \mathscr{A}_2(t_5, t_6)],$$

$$B_7^{(1)}(t_1, t_2, t_3, \ldots t_7) = \frac{1}{N^6}\Big[\mathcal{A}_7(t_1, t_2, \ldots t_7) - \underset{(21)}{\mathbf{S}}\,\mathcal{A}_2(t_1, t_2)\,\mathcal{A}_5(t_3, t_4, \ldots t_7)$$

$$- \underset{(35)}{\mathbf{S}}\,\mathcal{A}_3(t_1, t_2, t_3)\,\mathcal{A}_4(t_4, t_5, t_6, t_7) + 2!\,\underset{(105)}{\mathbf{S}}\,\mathcal{A}_2(t_1, t_2)\,\mathcal{A}_2(t_3, t_4)\,\mathcal{A}_3(t_5, t_6, t_7)\Big],$$

$$B_8^{(1)}(t_1, t_2, t_3, \ldots t_8) = \frac{1}{N^7}\Big[\mathcal{A}_8(t_1, t_2, \ldots t_8) - \underset{(28)}{\mathbf{S}}\,\mathcal{A}_2(t_1, t_2)\,\mathcal{A}_6(t_3, t_4, \ldots t_8)$$

$$- \underset{(56)}{\mathbf{S}}\,\mathcal{A}_3(t_1, t_2, t_3)\,\mathcal{A}_5(t_4, t_5, \ldots t_8) - \underset{(35)}{\mathbf{S}}\,\mathcal{A}_4(t_1, t_2, t_3, t_4)\,\mathcal{A}_4(t_5, t_6, t_7, t_8)$$

$$+ 2!\,\underset{(280)}{\mathbf{S}}\,\mathcal{A}_2(t_1, t_2)\,\mathcal{A}_3(t_3, t_4, t_5)\,\mathcal{A}_3(t_6, t_7, t_8)$$

$$+ 2!\,\underset{(210)}{\mathbf{S}}\,\mathcal{A}_2(t_1, t_2)\,\mathcal{A}_2(t_3, t_4)\,\mathcal{A}_4(t_5, t_6, t_7, t_8)$$

$$- 3!\,\underset{(105)}{\mathbf{S}}\,\mathcal{A}_2(t_1, t_2)\,\mathcal{A}_2(t_3, t_4)\,\mathcal{A}_2(t_5, t_6)\,\mathcal{A}_2(t_7, t_8)\Big] \quad \ldots\ldots\ldots(39).$$

I shall make one remark only concerning these formulae.

One might think that there is a difference between the right-hand sides of (38) and the series, such as (H^{ter}), which we have used above for obtaining the product moments of the $\delta_1 s_p'$'s. There is no difference, however, for the series (H^{ter}) involves terms which do not correspond to any product moment—such terms being eliminated in a natural way from $\mathcal{A}_4(t_1, t_2, t_3, t_4)$. Consequently it is immaterial which of these series we use for our calculations.

11. In the previous paragraph I have dealt with the correlated distribution of the $\delta_1 s_p'$'s. In order to obtain better approximations or to have a certain measure of the difference between the true values and those obtained by my previous method, further calculations concerning $\delta_2 s_p'$, $\delta_3 s_p'$, ... are required. The method which I am going to adopt is that which I indicated in the first section of this paper, and the point from which I shall start will still be the relation (8):

$$s_1'\frac{t}{1!} + s_2'\frac{t^2}{2!} + s_3'\frac{t^3}{3!} + \ldots + s_p'\frac{t^p}{p!} + \ldots$$

$$= \operatorname{Log}\left(1 + m_1'\frac{t}{1!} + m_2'\frac{t^2}{2!} + \ldots + m_q'\frac{t^q}{q!} + \ldots\right) \ldots\ldots\ldots(40).$$

If we put $\quad G_r(t) = \frac{t}{1!}\delta_r s_1' + \frac{t^2}{2!}\delta_r s_2' + \frac{t^3}{3!}\delta_r s_3' + \ldots + \frac{t^p}{p!}\delta_r s_p' + \ldots \quad \ldots\ldots(41),$

owing to the fact that $\delta_k m_p' = \delta m_p'$ the relation (40) gives

$$G_r(t) = G_1(t) - \frac{1}{2}(G_1(t))^2 + \frac{1}{3!}(G_1(t))^3 - \ldots + \frac{(-1)^{r-1}}{r}(G_1(t))^2 \quad \ldots\ldots(42).$$

This relation enables us to obtain any desired degree of approximation in the mean product moments of the sampling semi-invariants. And what is more, we shall not need new calculations, for any function associated with $G_r(t)$ will be expressed in terms of the functions associated with $G_1(t)$, functions with which we have already dealt and which are all worked out.

Let $A_p^{(r)}(t_1, t_2, \ldots t_p)$ be the pth order associated function with $G_r(t)$; we shall have:

$$A_p^{(r)}(t_1, t_2, \ldots t_p) = \int_{(D)} F G_r(t_1)\, G_r(t_2) \ldots G_r(t_p)\, dv$$

$$= \sum_1^r{}_{i_1} \sum_1^r{}_{i_2} \ldots \sum_1^r{}_{i_p} \frac{(-1)^{i_1+i_2+\ldots+i_p-p}}{i_1 \times i_2 \times \ldots \times i_p} A^{(1)}_{i_1+i_2+\ldots+i_p}(\underset{i_1}{t_1}, \underset{i_2}{t_2}, \ldots \underset{i_p}{t_p}) \quad \ldots\ldots\ldots(43),$$

where $A^{(1)}_{i_1+i_2+\ldots+i_p}(\underset{i_1}{t_1}, \underset{i_2}{t_2}, \ldots \underset{i_p}{t_p})$ stands for

$$A^{(i)}_{i_1+i_2+\ldots+i_p}(\underbrace{t_1, t_1, \ldots t_1}_{i_1}, \underbrace{t_2, t_2, \ldots t_2}_{i_2}, \ldots \underbrace{t_p, t_p, \ldots t_p}_{i_p}).$$

Before proceeding to calculate the approximated formulae for $A_p^{(r)}$ in terms of the characteristic function of the parent population, we observe that in order to obtain the exact values of $\delta s_p'$ we must make r equal to infinity in (42). Thus if

$$G(t) = \frac{t}{1!}\delta s_1' + \frac{t^2}{2!}\delta s_2' + \ldots + \frac{t^p}{p!}\delta s_p' + \ldots$$

we shall have

$$G(t) = \text{Log}\,[1 + G_1(t)] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(44).$$

By using the general formula (28), the function giving all the associated functions with the product moments of $\delta s_p'$ will be

$$\int_{(D)} F \times [1 + G_1(t_1)]^\lambda\, [1 + G_1(t_2)]^\lambda \ldots [1 + G_1(t_p)]^\lambda \ldots dv$$

$$= 1 + \frac{\lambda}{1!}\Sigma_i\, A_1'(t_i) + \frac{\lambda^2}{2!}\Sigma_{i,j}\, A_2'(t_i, t_j) + \ldots \quad \ldots\ldots\ldots\ldots\ldots(45);$$

hence

$$A_1'(t_1) = \sum_1^\infty{}_i \frac{(-1)^{i-1}}{i} A_i^{(1)}(\underset{i}{t_1}),$$

$$A_2'(t_1, t_2) = \sum_1^\infty{}_i \sum_1^\infty{}_j \frac{(-1)^{i+j-2}}{i \times j} A^{(2)}_{i+j}(\underset{i}{t_1}, \underset{j}{t_2}),$$

$$A_p'(t_1, t_2, \ldots t_p)$$

$$= \sum_1^\infty{}_{i_1} \sum_1^\infty{}_{i_2} \ldots \sum_1^\infty{}_{i_p} \frac{(-1)^{i_1+i_2+\ldots+i_p-p}}{i_1 \times i_2 \times \ldots \times i_p} A^{(1)}_{i_1+i_2+\ldots+i_p}(\underset{i_1}{t_1}, \underset{i_2}{t_2}, \ldots \underset{i_p}{t_p}) \quad \ldots\ldots(46).$$

Now owing to the formula (25) the associated functions with the semi-invariants are straightforward:

$$B_2(t_1, t_2) = \sum_1^\infty{}_i \sum_1^\infty{}_j \frac{(-1)^{i+j-2}}{i \times j} \left[A^{(1)}_{i+j}(\underset{i}{t_1}, \underset{j}{t_2}) - A_i^{(1)}(\underset{i}{t_1})\, A_j^{(1)}(\underset{j}{t_2}) \right],$$

$$B_3(t_1, t_2, t_3) = \sum_1^\infty{}_i \sum_1^\infty{}_j \sum_1^\infty{}_k \frac{(-1)^{i+j+k-3}}{i \times j \times k} \left[A^{(1)}_{i+j+k}(\underset{i}{t_1}, \underset{j}{t_2}, \underset{k}{t_3}) \right.$$

$$\left. - \mathbf{S} A_i^{(1)}(\underset{i}{t_1})\, A^{(i)}_{j+k}(\underset{j}{t_2}, \underset{k}{t_3}) + 2 A_i^{(1)}(\underset{i}{t_1})\, A_j^{(1)}(\underset{j}{t_2})\, A_k^{(1)}(\underset{k}{t_3}) \right],$$

$$B_p(t_1, t_2, \ldots t_p)$$
$$= \sum_1^{\infty}{}_{i_1} \sum_1^{\infty}{}_{i_2} \ldots \sum_1^{\infty}{}_{i_p} \frac{(-1)^{i_1+i_2+\cdots+i_p-p}}{i_1 \times i_2 \times \ldots \times i_p} \Big[A^{(1)}_{i_1+i_2+\ldots+i_p} (\underset{i_1}{t_1}, \underset{i_2}{t_2}, \ldots \underset{i_p}{t_p})$$
$$- \mathbf{SS}\, A^{(1)}_{i_1+i_2+\ldots+i_k} (\underset{i_1}{t_1}, \underset{i_2}{t_2}, \ldots \underset{i_k}{t_k})\, A^{(1)}_{i_{k+1}+\ldots+i_p} (\underset{i_{k+1}}{t_{k+1}}, \underset{i_{k+2}}{t_{k+2}}, \ldots \underset{i_p}{t_p})$$
$$+ 2!\,\mathbf{SS}\, A^{(1)}_{i_1+\ldots+i_k} (t_1, t_2, \ldots t_k)\, A^{(1)}_{i_{k+1}+\ldots+i_{k+l}} (t_{k+1}, \ldots t_{k+l})\, A^{(1)}_{i_{k+l+1}+\ldots+i_p} (\underset{i_{k+l+1}}{t_{k+l+1}}, \ldots t_p)$$
$$+ (-1)^{p-1} (p-1)!\, A^{(1)}_{i_1} (\underset{i_1}{t_1})\, A^{(1)}_{i_2} (\underset{i_2}{t_2}) \ldots A^{(1)}_{i_p} (\underset{i_p}{t_p}) \Big] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (47).$$

These formulae are very convenient especially for approximate calculations. Thus, in order to have an approximation as far as $\frac{1}{N^2}$, we must take in the development of the right-hand side of (47) only the terms of which the index does not exceed four.

Considering the expression of $A^{(1)}_p$ given above, we shall have

$$B_2(t_1, t_2) = \frac{1}{N} \mathcal{A}_2(t_1, t_2)$$
$$+ \frac{1}{N^2} [\mathcal{A}_2(\underset{2}{t_1}) \mathcal{A}_2(t_1, t_2) + \tfrac{1}{2}(\mathcal{A}_2(t_1, \underset{2}{t_2})) + \mathcal{A}_2(\underset{2}{t_2}) \mathcal{A}_2(t_1, t_2)$$
$$- \tfrac{1}{2} \mathcal{A}_3(\underset{2}{t_1}, t_2) - \tfrac{1}{2} \mathcal{A}_3(t_1, \underset{2}{t_2})] + \ldots,$$
$$B_3(t_1, t_2, t_3) = \frac{1}{N^2} [\mathcal{A}_3(t_1, t_2, t_3) - \mathcal{A}_2(t_1, t_2) \mathcal{A}_2(t_1, t_3)$$
$$- \mathcal{A}_2(t_2, t_1) \mathcal{A}_2(t_2, t_3) - \mathcal{A}_2(t_3, t_1) \mathcal{A}_2(t_3, t_2)] + \ldots,$$
$$B_4(t_1, t_2, t_3, t_4) = \frac{1}{N^2} [\mathcal{A}_4(t_1, t_2, t_3, t_4) + \mathcal{A}_2(t_1, t_2) \mathcal{A}_2(t_3, t_4)$$
$$+ \mathcal{A}_2(t_1, t_3) \mathcal{A}_2(t_2, t_4) + \mathcal{A}_2(t_1, t_4) \mathcal{A}_2(t_2, t_3)] + \ldots \quad \ldots\ldots\ldots\ldots (48).$$

If we compare these formulae with (39), it will be seen that in the development of $B_3(t_1, t_2, t_3)$ we obtained new terms in the coefficient of $\frac{1}{N^2}$. Although these terms are connected with the first approximation of B_3, they could not be obtained at that stage. The explanation is obvious: they follow from the fourth order product moment $\{\delta\varpi_i \delta\varpi_j \delta\varpi_k \delta\varpi_l\}$.

Substituting in (48) the expressions of $\mathcal{A}_p$ given by (37) we shall have

$$B_2(t_1, t_2) = \frac{1}{N} [e^{\phi(t_1+t_2)-\phi(t_1)-\phi(t_2)} - 1]$$
$$+ \frac{1}{N^2} [e^{\phi(t_1+t_2)+\phi(2t_1)-3\phi(t_1)-\phi(t_2)} - \tfrac{1}{2} e^{\phi(2t_1)-2\phi(t_1)}$$
$$+ e^{\phi(t_1+t_2)+\phi(2t_2)-\phi(t_1)-3\phi(t_2)} - \tfrac{1}{2} e^{\phi(2t_2)-2\phi(t_2)}$$
$$+ \tfrac{1}{2} e^{2\phi(t_1+t_2)-2\phi(t_1)-2\phi(t_2)} - e^{\phi(t_1+t_2)-\phi(t_1)-\phi(t_2)} + \tfrac{1}{2}$$
$$+ \tfrac{1}{2} e^{\phi(t_1+2t_2)-\phi(t_1)-2\phi(t_2)} - \tfrac{1}{2} e^{\phi(2t_1+t_2)-2\phi(t_1)-\phi(t_2)}] + \ldots,$$

$$B_3(t_1, t_2, t_3) = \frac{1}{N^2}\left[e^{\phi(t_1+t_2+t_3)-\phi(t_1)-\phi(t_2)-\phi(t_3)} - e^{\phi(t_1+t_2)+\phi(t_1+t_3)-2\phi(t_1)-\phi(t_2)-\phi(t_3)}\right.$$
$$- e^{\phi(t_1+t_2)+\phi(t_2+t_3)-\phi(t_1)-2\phi(t_2)-\phi(t_3)} - e^{\phi(t_1+t_3)+\phi(t_2+t_3)-\phi(t_1)-\phi(t_2)-2\phi(t_3)}$$
$$\left.+ e^{\phi(t_1+t_2)-\phi(t_1)-\phi(t_2)} + e^{\phi(t_2+t_3)-\phi(t_2)-\phi(t_3)} + e^{\phi(t_3+t_1)-\phi(t_3)-\phi(t_1)} - 1\right] + \ldots,$$

$$B_4(t_1, t_2, t_3, t_4) = \frac{1}{N^2}\left[\mathbf{S}\, e^{\phi(t_1+t_2)+\phi(t_3+t_4)-\phi(t_1)-\phi(t_2)-\phi(t_3)-\phi(t_4)}\right.$$
$$\left. - \mathbf{S}\, e^{\phi(t_1-t_2)-\phi(t_1)-\phi(t_2)} + 3\right] + \ldots \quad \ldots\ldots\ldots\ldots(49).$$

For practical calculations we can neglect in these formulae the terms which do not involve all the variates t—such as $e^{\phi(t_1+t_2)-\phi(t_1)-\phi(t_2)}$ in the development of $B_3(t_1, t_2, t_3)$—for these terms, as we noticed before, are reduced when the calculation is worked out.

To find the coefficient of $\frac{t_1^a}{a!}\cdot\frac{t_2^b}{b!}$ in the development of $B_2(t_1, t_2)$ we proceed as follows.

Let us consider first the series

$$e^{\phi(t_1+t_2)-\phi(t_1)-\phi(t_2)} = \sum_0^\infty{}_a \sum_0^\infty{}_b N_{a,b}^{(0)} \frac{t_1^a}{a!}\cdot\frac{t_2^b}{b!}.$$

Owing to the fact that

$$\phi(t_1+t_2) - \phi(t_1) - \phi(t_2)$$
$$= s_2\frac{t_1}{1!}\cdot\frac{t_2}{1!} + s_3\left(\frac{t_1^2}{2!}\cdot\frac{t_2}{1!} + \frac{t_1}{1!}\cdot\frac{t_2^2}{2!}\right) + s_4\left(\frac{t_1^3}{3!}\cdot\frac{t_2}{1!} + \frac{t_1^2}{2!}\cdot\frac{t_2^2}{2!} + \frac{t_1}{1!}\cdot\frac{t_2^3}{3!}\right) + \ldots,$$

we shall have

$$N_{a,b}^{(0)} = \mathbf{S}\frac{a!\,b!}{(i_1!\,j_1!)^{g_1}(i_2!\,j_2!)^{g_2}\ldots} \times \frac{1}{g_1!\,g_2!\ldots}\, s_{i_1+j_1}^{g_1}\, s_{i_2+j_2}^{g_2}\ldots \quad \ldots\ldots\ldots(50),$$

the sum being extended to all the values for which

$$\Sigma ig = a, \qquad \Sigma jg = b,$$

the i's and j's being always different from 0. This formula is similar to that which gives the product moments of a distribution of two variates in terms of the semi-invariants.

Continuing,

$$N_{a,b,c}^{(0)} = \mathbf{S}\frac{a!\,b!\,c!}{(i_1!\,j_1!\,k_1!)^{g_1}(i_2!\,j_2!\,k_2!)^{g_2}\ldots} \times \frac{1}{g_1!\,g_2!\ldots}\, s_{i_1+j_1+k_1}^{g_1}\, s_{i_2+j_2+k_2}^{g_2}\ldots \quad \ldots(51),$$

in which, as before,

$$\Sigma ig = a, \qquad \Sigma jg = b, \qquad \Sigma kg = c,$$
$$i \neq 0, \qquad j \neq 0, \qquad k \neq 0.$$

In the same way we obtain for the coefficient of $\frac{t_1^a}{a!}\cdot\frac{t_2^b}{b!}$ in

$$e^{\phi(t_1+t_2)+\phi(2t_1)-3\phi(t_1)-\phi(t_2)}$$

the formula

$$\mathbf{S}\frac{a!\,b!}{(i_1!\,j_1!)^{g_1}(i_2!\,j_2!)^{g_2}\ldots} \times \frac{1}{g_1!\,g_2!\ldots}\, s_{i_1,j_1}^{g_1}\, s_{i_2,j_2}^{g_2}\ldots \quad \ldots\ldots\ldots\ldots(52),$$

in which
$$\Sigma ig = a, \qquad \Sigma jg = b, \qquad i \neq 0,$$
and
$$\left.\begin{aligned} s_{i,j} &= s_{i+j} & \text{if } j \neq 0 \\ s_{i,j} &= (2^i - 2)\, s_i & \text{if } j = 0 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(53).$$

The coefficient of the same term in the development of
$$e^{\phi(2t_1+t_2)-2\phi(t_1)-\phi(t_2)}$$
will be given by the same formula (52), but with this difference, that the sum is extended to the values for which
$$\Sigma ig = a, \qquad \Sigma jg = b, \quad i \neq 0,$$
and
$$\left.\begin{aligned} s_{i,j} &= 2^i s_{i+j} & \text{if } j \neq 0 \\ s_{i,j} &= (2^i - 2)\, s_i & \text{if } j = 3 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(54).$$

The similarity of the previous formulae enables us to write down the following relation:
$$\begin{aligned} L_{a,b} = {} & \mathbf{S}\left(\frac{1}{N} - \frac{2^{a-1} + 2^{b-1} - 2^{g-1} - 1}{N^2}\right) \frac{a!\, b!}{(i_1!\, j_1!)^{g_1} (i_2!\, j_2!)^{g_2} \ldots} \times \frac{1}{g_1!\, g_2! \ldots} s_{i_1+j_1}^{g_1} s_{i_2+j_2}^{g_2} \ldots \\ & + \frac{1}{N^2} \mathbf{S}_i \frac{a!\, b!}{(i_1!\, j_1!)^{g_1} (i_2!\, j_2!)^{g_2} \ldots} \times \frac{1}{g_1!\, g_2! \ldots} s_{i_1,j_1}^{g_1} s_{i_2,j_2}^{g_2} \ldots \\ & + \frac{1}{N^2} \mathbf{S}_j \frac{a!\, b!}{(i_1!\, j_1!)^{g_1} (i_2!\, j_2!)^{g_2} \ldots} \times \frac{1}{g_1!\, g_2! \ldots} s_{i_1,j_1}^{g_1} s_{i_2,j_2}^{g_2} \ldots \\ & - \frac{1}{2N^2} \mathbf{S}_i' \frac{a!\, b!}{(i_1!\, j_1!)^{g_1} (i_2!\, j_2!)^{g_2} \ldots} \times \frac{1}{g_1!\, g_2! \ldots} s_{i_1,j_1}^{g_1} s_{i_2,j_2}^{g_2} \ldots \\ & - \frac{1}{2N^2} \mathbf{S}_j' \frac{a!\, b!}{(i_1!\, j_1!)^{g_1} (i_2!\, j_2!)^{g_2} \ldots} \times \frac{1}{g_1!\, g_2! \ldots} s_{i_1,j_1}^{g_1} s_{i_2,j_2}^{g_2} \ldots \quad \ldots\ldots\ldots\ldots(55). \end{aligned}$$

In this formula the sums are extended to all the values for which
$$\Sigma ig = a, \qquad \Sigma jg = b, \qquad \Sigma g_k = g,$$
the sum $\mathbf{S}$ involving the values of i and j different from 0, whereas in the sum $\mathbf{S}_i$ at least one of the values of i is equal to 0, in which case
$$\begin{aligned} s_{i,j} &= s_{i+j} & \text{if } & i \neq 0, j \neq 0, \\ s_{i,j} &= (2^j - 2)\, s_j & \text{if } & i = 0, j \neq 0, \end{aligned}$$
for $\mathbf{S}_i$, and
$$\begin{aligned} s_{i,j} &= 2^j s_{i+j} & \text{if } & i \neq 0, j \neq 0, \\ s_{i,j} &= (2^j - 2)\, s_j & \text{if } & i = 0, j \neq 0, \end{aligned}$$
for $\mathbf{S}_i'$.

Similar formulae may be easily established for the other developments, but it is not necessary to dwell on them now; the reader will not have any difficulty in writing them down.

However, I have calculated a table of the first coefficients of the development of $B_2(t_1, t_2)$—i.e. of the first second order product moments of the sampling semi-invariants—which will be found at the end of this paper.

Associated Functions of the Distribution of Sampling Product Moments about the Mean.

12. In this section I shall make an application of the method of associated functions to the only problem which has not yet been attacked in the previous pages, to the distribution of the sampling product moments calculated about the mean of the sample.

If
$$\mu_r' = \frac{1}{N}\sum_1^N{}_i (x_i - m_1')^r,$$
we shall have
$$\chi(t) = 1 + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \ldots + \mu_p' \frac{t_p}{p!} + \ldots$$
$$= \left(1 + m_1' \frac{t}{1!} + m_2' \frac{t^2}{2!} + \ldots + m_q' \frac{t_q}{q!} + \ldots\right) e^{-m_1't} \quad \text{.........(56).}$$

Let us put
$$\gamma_r(t) = \frac{t^2}{2!}\delta_r\mu_2' + \frac{t^3}{3!}\delta_r\mu_3' + \ldots + \frac{t^p}{p!}\delta_r\mu_p' + \ldots \quad \text{.........(57);}$$
since
$$\frac{\partial^i\chi}{\partial m_1'^i} = (-t)^i e^{-m_1t}\left(1 + m_1' \frac{t}{1!} + m_2' \frac{t^2}{2!} + \ldots + m_p' \frac{t^p}{p!} + \ldots\right) + i(-1)^{i-1}t^i e^{-m_1't},$$
$$\frac{\partial^i\chi}{\partial m_1'^{i-1}\partial m_k'} = (-t)^{i-1}\frac{t^k}{k!}e^{-m_1't},$$
$$\frac{\partial^i\chi}{\partial m_1'^{i-2}\partial m_k'\partial m_l'} - 0,$$
we shall have
$$\gamma_r(t) = g(t)\left[1 - \frac{t}{1!}\delta m_1' + \frac{t^2}{2!}\delta^2 m_1' - \ldots + \frac{(-t)^{r-1}}{(r-1)!}\delta^{r-1}m_1'\right]$$
$$- e^{\phi(t)}\left[\frac{t}{1!}\delta m_1' - \frac{t^2}{2!}\delta^2 m_1' + \ldots - \frac{(-t)^r}{r!}\delta^r m_1'\right] \quad \text{......(58),}$$
where we have supposed that the origin is taken at the mean of the parent population.

If $r = l$ we obtain $\qquad \delta_1\mu_p' = \delta m_p' - p\delta m'_{p-1}\delta m_1',$

a formula given by Professor Karl Pearson*.

Now let us make in (58) r equal to infinity; we shall have
$$\gamma(t) = e^{-m_1't}[g(t) + e^{\phi(t)}] - e^{\phi(t)} \quad \text{......................(59).}$$

Substituting in this relation the value of $g(t)$ given by (29) we obtain
$$\gamma(t) = \frac{1}{N}\sum_1^N{}_i e^{\frac{t}{N}\sum_1^N{}_k (x_i - x_k)} - e^{\phi(t)} \quad \text{......................(60).}$$

* *Biometrika*, Editorials, Vol. II. pp. 273—281 and Vol. IX. pp. 1—10.

Before commencing the calculation of the associated functions of $\gamma(t)$ I shall consider the expression

$$\Omega_p(t_1, t_2, \dots t_p) = \frac{1}{N^p}\int_{(D)} F \sum_{i_1}^{N} {}_1 \sum_{i_2}^{N} {}_1 \dots \sum_{i_p}^{N} {}_1 \, e^{\frac{1}{N}\sum_j^p {}_1 \sum_{k_j}^N {}_1 t_j (x_{i_j} - x_{k_j})} \, dv$$

$$= e^{N\phi\left(-\frac{T}{N}\right) + \Sigma_i \phi_i} \left[\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{p-1}{N}\right)\right.$$

$$+ \frac{1}{N}\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{p-2}{N}\right)(\mathbf{S} e^{\phi_{12}})$$

$$+ \frac{1}{N^2}\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{p-3}{N}\right)(\mathbf{S} e^{\phi_{123}} + \mathbf{S} e^{\phi_{12}+\phi_{34}})$$

$$+ \frac{1}{N^3}\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{p-4}{N}\right)(\mathbf{S} e^{\phi_{1234}} + \mathbf{S} e^{\phi_{12}+\phi_{345}} + \mathbf{S} e^{\phi_{12}+\phi_{34}+\phi_{56}})$$

$$+ \frac{1}{N^4}\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{p-5}{N}\right)(\mathbf{S} e^{\phi_{12345}} + \mathbf{S} e^{\phi_{12}+\phi_{3456}} + \mathbf{S} e^{\phi_{123}+\phi_{456}}$$

$$+ \mathbf{S} e^{\phi_{12}+\phi_{34}+\phi_{567}} + \mathbf{S} e^{\phi_{12}+\phi_{34}+\phi_{56}+\phi_{78}})$$

$$\left. + \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \right] \quad \dots\dots(61),$$

where $$T = t_1 + t_2 + \dots + t_p,$$

and

$$\phi_{i_1 i_2 \dots i_k} = \phi\left(t_{i_1} + t_{i_2} + \dots + t_{i_k} - \frac{T}{N}\right) - \phi\left(t_{i_1} - \frac{T}{N}\right) - \phi\left(t_{i_2} - \frac{T}{N}\right) - \dots$$

$$- \phi\left(t_{i_k} - \frac{T}{N}\right) + (k-1)\,\phi\left(-\frac{T}{N}\right).$$

We observe that owing to the relation (28), in calculating the associated functions of the distribution of the μ's, we can neglect the second term on the right-hand side of (60)—i.e. $e^{\phi(t)}$—for that amounts to a "*change of origin.*" Thus

$$\beta_p(t_1, t_2, \dots t_p) = \Omega_p(t_1, t_2, \dots t_p) - \mathbf{SS}\,\Omega_k(t_1, t_2, \dots t_k)\,\Omega_{p-k}(t_{k+1}, t_{k+2}, \dots t_p)$$

$$+ 2!\,\mathbf{SS}\,\Omega_k(t_1, t_2, \dots t_k)\,\Omega_l(t_{k+1}, \dots t_{k+l})\,\Omega_{p-k-l}(t_{k+l+1}, \dots t_p)$$

$$\dots\dots(62);$$

hence

$$\beta_2(t_1, t_2) = \left(1 - \frac{1}{N}\right) e^{(N-2)\phi\left(-\frac{t_1+t_2}{N}\right) + \phi\left(t_1 - \frac{t_1+t_2}{N}\right) + \phi\left(t_2 - \frac{t_1+t_2}{N}\right)}$$

$$+ \frac{1}{N} e^{(N-1)\phi\left(-\frac{t_1+t_2}{N}\right) + \phi\left(t_1 + t_2 - \frac{t_1+t_2}{N}\right)}$$

$$- e^{(N-1)\left[\phi\left(-\frac{t_1}{N}\right) + \phi\left(-\frac{t_2}{N}\right)\right] + \phi\left(t_1 - \frac{t_1}{N}\right) + \phi\left(t_2 - \frac{t_2}{N}\right)} \quad \dots\dots(63),$$

$$\beta_3(t_1, t_2, t_3) = \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)e^{(N-3)\phi\left(-\frac{T}{N}\right)+\phi\left(t_1-\frac{T}{N}\right)+\phi\left(t_2-\frac{T}{N}\right)+\phi\left(t_3-\frac{T}{N}\right)}$$

$$+\frac{1}{N}\left(1-\frac{1}{N}\right)\mathbf{S}\, e^{(N-2)\phi\left(-\frac{T}{N}\right)+\phi\left(t_1-\frac{T}{N}\right)+\phi\left(t_2+t_3-\frac{T}{N}\right)}$$

$$+\frac{1}{N^2}e^{(N-1)\phi\left(-\frac{T}{N}\right)+\phi\left(T-\frac{T}{N}\right)}$$

$$-\left(1-\frac{1}{N}\right)\mathbf{S}\, e^{(N-2)\phi\left(-\frac{t_1+t_2}{N}\right)+\phi\left(t_1-\frac{t_1+t_2}{N}\right)+\phi\left(t_2-\frac{t_1+t_2}{N}\right)+(N-1)\phi\left(-\frac{t_3}{N}\right)+\phi\left(t_3-\frac{t_3}{N}\right)}$$

$$-\frac{1}{N}\mathbf{S}\, e^{(N-1)\phi\left(-\frac{t_1+t_2}{N}\right)+\phi\left(t_1+t_2-\frac{t_1+t_2}{N}\right)+(N-1)\phi\left(-\frac{t_3}{N}\right)+\phi\left(t_3-\frac{t_3}{N}\right)}$$

$$+2e^{(N-1)\left[\phi\left(-\frac{t_1}{N}\right)+\phi\left(-\frac{t_2}{N}\right)+\phi\left(-\frac{t_3}{N}\right)\right]+\phi\left(t_1-\frac{t_1}{N}\right)+\phi\left(t_2-\frac{t_2}{N}\right)+\phi\left(t_3-\frac{t_3}{N}\right)} \quad \text{......(64).}$$

13. In order to deduce some practical results from (62), I shall use—as I did before for the distribution of the sampling semi-invariants—the well-known formula which gives the moments in terms of the semi-invariants.

Let us consider the characteristic function of a multivariate distribution

$$\sum_0^\infty{}_a \sum_0^\infty{}_b \sum_0^\infty{}_c \ldots m_{a,b,c,\ldots}\frac{t^a}{a!}\cdot\frac{u^b}{b!}\cdot\frac{v^c}{c!}\ldots = e^{\sum_0{}_\alpha' \sum_0{}_\beta' \sum_0{}_\gamma' \ldots s_{\alpha,\beta,\gamma,\ldots}\frac{t^\alpha}{\alpha!}\cdot\frac{u^\beta}{\beta!}\cdot\frac{v^\gamma}{\gamma!}\ldots}.$$

We shall have

$$m_{a,b,c,\ldots} = \mathbf{S}\,\frac{a!\,b!\,c!}{(i_1!\,j_1!\,k_1!\ldots)^{g_1}(i_2!\,j_2!\,k_2!\ldots)^{g_2}\ldots}\times\frac{1}{g_1!\,g_2!\ldots}\,s^{g_1}_{i_1,j_1,k_1,\ldots}\,s^{g_2}_{i_2,j_2,k_2\ldots} \quad \text{.........(65),}$$

where the sum is extended to all the multipartitions

$$\Sigma ig = a, \quad \Sigma jg = b, \quad \Sigma kg = c, \ldots.$$

The developments of the terms involved by the formula (63) will be

$$(N-2)\phi\left(-\frac{t_1+t_2}{N}\right)+\phi\left(t_1-\frac{t_1+t_2}{N}\right)+\phi\left(t_2-\frac{t_1+t_2}{N}\right)$$

$$=\sum_0^\infty{}_\alpha'\sum_0^\infty{}_\beta'\left(-\frac{1}{N}\right)^{\alpha+\beta}[(1-N)^\alpha+(1-N)^\beta+N-2]\,s_{\alpha+\beta}\frac{t_1^\alpha}{\alpha!}\cdot\frac{t_2^\beta}{\beta!},$$

$$(N-1)\phi\left(-\frac{t_1+t_2}{N}\right)+\phi\left(t_1+t_2-\frac{t_1+t_2}{N}\right)$$

$$=\sum_0^\infty{}_\alpha'\sum_0^\infty{}_\beta'\left(-\frac{1}{N}\right)^{\alpha+\beta}[(1-N)^{\alpha+\beta}-(1-N)]\,s_{\alpha+\beta}\frac{t_1^\alpha}{\alpha!}\cdot\frac{t_2^\beta}{\beta!},$$

$$(N-1)\phi\left(-\frac{t_1}{N}\right)+\phi\left(t_1-\frac{t_1}{N}\right)$$

$$=\sum_1^\infty{}_\alpha\left(-\frac{1}{N}\right)^\alpha[(1-N)^\alpha-(1-N)]\,s_\alpha\frac{t_1^\alpha}{\alpha!}, \quad \text{......................(66).}$$

Now, if we designate by $S(a, b)$ the second order semi-invariant corresponding to μ_a' and μ_b', by using the formula (65) we shall obtain

$$\begin{aligned} S(a, b) = {} & \left(-\frac{1}{N}\right)^{a+b}\left(1-\frac{1}{N}\right) \mathbf{S} \frac{a\,!\,b\,!}{(i_1\,!\,j_1\,!)^{g_1}(i_2\,!\,j_2\,!)^{g_2}\ldots} \times \frac{1}{g_1\,!\,g_2\,!\ldots} \\ & \times [N-2+(1-N)^{i_1}+(1-N)^{j_1}]^{g_1}[N-2+(1-N)^{i_2}+(1-N)^{j_2}]^{g_2}\ldots \\ & \qquad \times s_{i_1+j_1}^{g_1}\, s_{i_2+j_2}^{g_2}\ldots \\ & + (-1)^{a+b}\frac{1}{N^{a+b+1}} \mathbf{S} \frac{a\,!\,b\,!}{(i_1\,!\,j_1\,!)^{g_1}(i_2\,!\,j_2\,!)^{g_2}\ldots} \times \frac{1}{g_1\,!\,g_2\,!\ldots} \\ & \times [(1-N)^{i_1+j_1}-(1-N)]^{g_1}[(1-N)^{i_2+j_2}-(1-N)]^{g_2}\ldots \times s_{i_1+j_1}^{g_1}\, s_{i_2+j_2}^{g_2}\ldots \\ & - \left(-\frac{1}{N}\right)^{a+b} \mathbf{S} \frac{a\,!}{(i_1\,!)^{g_1}(i_2\,!)^{g_2}\ldots} \times \frac{1}{g_1\,!\,g_2\,!\ldots}[(1-N)^{i_1}-(1-N)]^{g_1}\ldots \times s_{i_1}^{g_1}\, s_{i_2}^{g_2}\ldots \\ & \times \mathbf{S} \frac{b\,!}{(j_1\,!)^{g_1'}(j_2\,!)^{g_2'}\ldots} \times \frac{1}{g_1'\,!\,g_2'\,!\ldots}[(1-N)^{j_1}-(1-N)]^{g_1'}\ldots \times s_{j_1}^{g_1'}\, s_{j_2}^{g_2'}\ldots \end{aligned} \qquad \ldots\ldots\ldots(67).$$

I shall make an application of this formula to the calculation of $S(2, 4)$. In that case we shall have the following bipartitions:

$$\begin{array}{c|c} 1 & \\ \hline 2 & 2 \\ 4 & 4 \\ \hline 6 & 6 \end{array} \qquad \text{giving } s_6,$$

$$\begin{array}{cc|c} 1 & 1 & \\ \hline 2 & \cdot & 2 \\ 2 & 2 & 4 \\ \hline 4 & 2 & 6 \end{array} \qquad \begin{array}{cc|c} 1 & 1 & \\ \hline 1 & 1 & 2 \\ 3 & 1 & 4 \\ \hline 4 & 2 & 6 \end{array} \qquad \begin{array}{cc|c} 1 & 1 & \\ \hline \cdot & 2 & 2 \\ 4 & \cdot & 4 \\ \hline 4 & 2 & 6 \end{array} \qquad \text{giving } s_4 s_2,$$

$$\begin{array}{cc|c} 1 & 1 & \\ \hline 2 & \cdot & 2 \\ 1 & 3 & 4 \\ \hline 3 & 3 & 6 \end{array} \qquad \begin{array}{c|c} 2 & \\ \hline 1 & 2 \\ 2 & 4 \\ \hline 3 & 6 \end{array} \qquad \text{giving } s_3^2,$$

$$\begin{array}{cc|c} 2 & 1 & \\ \hline 1 & \cdot & 2 \\ 1 & 2 & 4 \\ \hline 2 & 2 & 6 \end{array} \qquad \begin{array}{cc|c} 2 & 1 & \\ \hline \cdot & 2 & 2 \\ 2 & \cdot & 4 \\ \hline 2 & 2 & 6 \end{array} \qquad \text{giving } s_2^3.$$

We observe that not all these bipartitions must be considered for the calculation which is to follow. Thus, the last bipartitions of those giving $s_4 s_2$ and s_2^3 must be neglected, for the expressions following from them vanish. It is easy to see that, for these bipartitions, the terms which proceed from the first two sums of (67) are reduced by those of the double sum of the same relation. Therefore we must neglect at the same time the terms following from this double sum.

The terms corresponding to each bipartition are now calculated in turn. The coefficient of s_6 is

$$\frac{1}{N^6}\left(1-\frac{1}{N}\right)[N-2+(1-N)^4+(1-N)^2]+\frac{1}{N^7}[(1-N)^6-(1-N)]$$
$$=\frac{1}{N}\left(1-\frac{1}{N}\right)^2\left(1-\frac{3}{N}+\frac{3}{N^2}\right).$$

The coefficient of s_4s_2 is

$$\frac{6}{N^6}\left(1-\frac{1}{N}\right)[N-2+(1-N)^2+1][N-2+2(1-N)^2]$$
$$+\frac{8}{N^6}\left(1-\frac{1}{N}\right)[N-2+(1-N)^3+1][N-2+2(1-N)]$$
$$+\frac{14}{N^7}[(1-N)^4-(1-N)][(1-N)^2-(1-N)]=\frac{2}{N}\left(1-\frac{1}{N}\right)\left(7-\frac{18}{N}+\frac{15}{N^2}\right).$$

In the same manner the coefficients of s_3^2 and s_2^3 are found to be respectively

$$\frac{6}{N}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)^2,\qquad \frac{12}{N}\left(1-\frac{1}{N}\right)^2;$$

hence

$$S(2,4)=\frac{1}{N}\left(1-\frac{1}{N}\right)^2\left(1-\frac{3}{N}+\frac{3}{N^2}\right)s_6$$
$$+\frac{2}{N}\left(1-\frac{1}{N}\right)\left(7-\frac{18}{N}+\frac{15}{N^2}\right)s_4s_2$$
$$+\frac{6}{N}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)^2s_3^2+\frac{12}{N}\left(1-\frac{1}{N}\right)^2s_2^3\;\ldots\ldots\ldots\ldots(68).$$

Craig has obtained this formula by a different method*.

By a closer study of the formulae (61) and (62) one may construct a general procedure for obtaining any semi-invariant of the correlated distribution of sampling moments. This method is based upon the calculus of multi-partitions, and personally I think that this is a natural way of attacking similar problems, the multi-partitions being necessarily introduced into our calculations by the form of the formula giving the semi-invariants in terms of the moments or *vice versa*†.

The rules which enable us to make use of the calculus of multi-partitions are the following:

(1) The calculations will be made easier if we suppose that we deal with a sample of $N+1$.

(2) We shall eliminate from our calculations all the multi-partitions which have not all the rows connected by elements of columns such as

$$\begin{matrix}\cdot&7&2&\cdot\\ \cdot&1&5&\cdot\\ 3&\cdot&\cdot&6\\ \cdot&4&2&\cdot\end{matrix}\qquad\qquad \begin{matrix}2&\cdot&\cdot&\cdot\\ 4&2&\cdot&\cdot\\ \cdot&\cdot&3&3\\ \cdot&\cdot&4&5\end{matrix}$$

* See C. C. Craig, *loc. cit.* p. 32.

† A similar procedure, but for the cumulative moments of the sampling cumulative moments, has been indicated by R. A. Fisher. The considerations which led him to the result are, however, different from those contained in this paper.

We observe that these multi-partitions may be decomposed into several others.

(3) The coefficient derived from a p row partition will be a fraction. The denominator of this fraction will be $(-1)^s(N+1)^{s+p-1}$, s being the arithmetical sum of the elements of the corresponding partition.

(4) To obtain the corresponding numerator we shall proceed as follows:

(*a*) We shall form all the multi-partitions which may be derived from the initial one by adding in every possible way the elements of the columns. Thus the multi-partitions derived from the four row partition

$$\begin{array}{cccc} 1 & \cdot & 2 & \cdot \\ 1 & 1 & \cdot & 1 \\ \cdot & 2 & 1 & 1 \\ 1 & 1 & \cdot & 2 \end{array}$$

will be

$$\left.\begin{array}{cccc} 2&1&2&1\\ \cdot&2&1&1\\ 1&1&\cdot&2 \end{array}\quad \begin{array}{cccc} 1&2&3&1\\ 1&1&\cdot&1\\ 1&1&\cdot&2 \end{array}\quad \begin{array}{cccc} 2&1&2&2\\ 1&1&\cdot&1\\ \cdot&2&1&1 \end{array}\quad \begin{array}{cccc} 1&\cdot&2&\cdot\\ 1&3&1&2\\ 1&1&\cdot&2 \end{array}\quad \begin{array}{cccc} 1&\cdot&2&\cdot\\ 2&2&3&\cdot\\ \cdot&2&1&1 \end{array}\quad \begin{array}{cccc} 1&\cdot&2&\cdot\\ 1&1&\cdot&1\\ 1&3&1&3 \end{array}\right\}\text{ 3 rows,}$$

$$\left.\begin{array}{cccc} 2&1&2&1\\ 1&3&1&3 \end{array}\quad \begin{array}{cccc} 1&2&3&1\\ 2&2&\cdot&3 \end{array}\quad \begin{array}{cccc} 2&1&2&2\\ 1&3&1&2 \end{array}\quad \begin{array}{cccc} 2&3&3&2\\ 1&1&\cdot&2 \end{array}\quad \begin{array}{cccc} 3&2&2&3\\ \cdot&2&1&1 \end{array}\quad \begin{array}{cccc} 2&3&3&3\\ 1&1&\cdot&1 \end{array}\quad \begin{array}{cccc} 2&4&1&4\\ 1&\cdot&2&\cdot \end{array}\right\}\text{ 2 rows,}$$

$$\left.\begin{array}{cccc} 3&4&3&4 \end{array}\right\}\text{ 1 row.}$$

(*b*) A p row partition will contribute with the first factor

$$N(N-1)(N-2)\dots(N-p+2),$$

which must be considered equal to 1 in the case of $p=1$.

(*c*) Each column contributes with the factor

$$N-p+1+(-N)^{a_1}+(-N)^{a_2}+\dots+(-N)^{a_p},$$

where $a_1, a_2, \dots a_p$ are the elements of the column considered.

(*d*) Finally the numerical coefficient will be given by the well-known expression

$$\frac{a!\,b!\,c!\dots}{(i_1!\,j_1!\dots)^{g_1}(i_2!\,j_2!\dots)^{g_2}}\times\frac{1}{g_1!\,g_2!\dots}.$$

Although the calculation of any coefficient can be carried out through these rules it is preferable to obtain some new results in order to shorten the labour.

Let us denote by

$$p\left|\begin{array}{lllll} a_1^1 & a_1^2 & a_1^3 & \dots & a_1^n \\ a_2^1 & a_2^2 & a_2^3 & \dots & a_2^n \\ \dots & \dots & \dots & \dots & \dots \\ a_m^1 & a_m^2 & a_m^3 & \dots & a_m^n \end{array}\right\rfloor$$

the numerator of the function of N connected with the multi-partition between brackets and with those derived from it. On the other hand let us put

$$\begin{Vmatrix} a_1^1 & a_1^2 & \dots & a_1^n \\ a_2^1 & a_2^2 & \dots & a_2^n \\ \dots & \dots & \dots & \dots \\ a_m^1 & a_m^2 & \dots & a_m^n \end{Vmatrix} = \prod_1^n [N - m + 1 + (-N)^{a_1^i} + (-N)^{a_2^i} + \dots + (-N)^{a_m^i}];$$

we shall have

$$P\begin{bmatrix} a_1^1 & a_1^2 & \dots & a_1^n \\ a_2^1 & a_2^2 & \dots & a_2^n \\ \dots & \dots & \dots & \dots \\ a_m^1 & a_m^2 & \dots & a_m^n \end{bmatrix} = N(N-1)\dots(N-m+2)\begin{Vmatrix} a_1^1 & a_1^2 & \dots & a_1^n \\ a_2^1 & a_2^2 & \dots & a_2^n \\ \dots & \dots & \dots & \dots \\ a_m^1 & a_m^2 & \dots & a_m^n \end{Vmatrix}$$

$$+ N(N-1)\dots(N-m+3)\,\mathbf{S}\begin{Vmatrix} a_1^1 + a_2^1 & a_1^2 + a_2^2 & \dots & a_1^n + a_2^n \\ a_3^1 & a_3^2 & \dots & a_3^n \\ \dots & \dots & \dots & \dots \\ a_m^1 & a_m^2 & \dots & a_m^n \end{Vmatrix}$$

$$+ N(N-1)\dots(N-m+4)\,\mathbf{S}\begin{Vmatrix} a_1^1 + a_2^1 + a_3^1 & a_1^2 + a_2^2 + a_3^2 & \dots & a_1^n + a_2^n + a_3^n \\ a_4^1 & a_4^2 & \dots & a_4^n \\ \dots & \dots & \dots & \dots \\ a_m^1 & a_m^2 & \dots & a_m^n \end{Vmatrix}$$

$$+ N(N-1)\dots(N-m+4)\,\mathbf{S}\begin{Vmatrix} a_1^1 + a_2^1 & a_1^2 + a_2^2 & \dots & a_1^n + a_2^n \\ a_3^1 + a_4^1 & a_3^2 + a_4^2 & \dots & a_3^n + a_4^n \\ a_5^1 & a_5^2 & \dots & a_5^n \\ \dots & \dots & \dots & \dots \\ a_m^1 & a_m^2 & \dots & a_m^n \end{Vmatrix}$$

$$+ \dots \dots\dots\dots\dots\dots\dots\dots\dots\dots (a).$$

It is not at all difficult to show now that, if $m = n$, we have

$$P\begin{bmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 \\ \dots & \dots & \dots & \dots \\ a_1^n & a_2^n & \dots & a_n^n \end{bmatrix} = P\begin{bmatrix} a_1^1 & a_1^2 & \dots & a_1^n \\ a_2^1 & a_2^2 & \dots & a_2^n \\ \dots & \dots & \dots & \dots \\ a_n^1 & a_n^2 & \dots & a_n^n \end{bmatrix} \dots\dots\dots\dots (b),$$

and that

$$P\begin{bmatrix} a_1^1 & a_1^2 & \dots & a_1^n & 0 & 0 & \dots & 0 \\ a_2^1 & a_2^2 & \dots & a_2^n & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_m^1 & a_m^2 & \dots & a_m^n & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \underbrace{0 & 0 & \dots & 0}_{\beta} \end{bmatrix} \left.\vphantom{\begin{matrix}0\\ \dots\\ 0\end{matrix}}\right\}\alpha = (N+1)^{\alpha+\beta}\, P\begin{bmatrix} a_1^1 & a_1^2 & \dots & a_1^n \\ a_2^1 & a_2^2 & \dots & a_2^n \\ \dots & \dots & \dots & \dots \\ a_m^1 & a_m^2 & \dots & a_m^n \end{bmatrix};$$

since we deduce that we can change the columns into rows and *vice versa*, the function P will not be altered even if $n = m$.

Another important property of the function P is the following:

$$P\begin{bmatrix} a_1^1 & a_1^2 & \dots & a_1^n & \lambda_1 \\ a_2^1 & a_2^2 & \dots & a_2^n & \lambda_2 \\ \\ a_m^1 & a_m^2 & \dots & a_m^n & \lambda_m \end{bmatrix}$$

$$= [N - m + 1 + (-N)^{\lambda_1} + (-N)^{\lambda_2} + \dots + (-N)^{\lambda_m}]\, P\begin{bmatrix} a_1^1 & a_1^2 & \dots & a_1^n \\ a_2^1 & a_2^2 \\ \\ a_m^1 & a_m^2 & \dots & a_m^n \end{bmatrix}$$

$$+ \mathbf{S}[(-N)^{\lambda_1} - 1][(-N)^{\lambda_2} - 1]\, P\begin{bmatrix} a_1^1 + a_2^1 & a_1^2 + a_2^2 & \dots & a_1^n + a_2^n \\ a_3^1 & a_3^2 & \dots & a_3^n \\ \\ a_m^1 & a_m^2 & \dots & a_m^n \end{bmatrix}$$

$$+ \mathbf{S}[(-N)^{\lambda_1} - 1][(-N)^{\lambda_2} - 1][(-N)^{\lambda_3} - 1]$$

$$\times P\begin{bmatrix} a_1^1 + a_2^1 + a_3^1 & a_1^2 + a_2^2 + a_3^2 & \dots & a_1^n + a_2^n + a_3^n \\ a_4^1 & a_4^2 & \dots & a_4^n \\ \dots & \dots & \dots & \dots \\ a_m^1 & a_m^2 & \dots & a_m^n \end{bmatrix}$$

$$+ \dots \qquad \dots\dots\dots\dots\dots\dots\dots\dots \quad (c).$$

As we have

$$P[a_1 + 1, a_2 + 1, \dots a_n + 1] = [(-N)^{a_1+1} + N][(-N)^{a_2+1} + N] \dots [(-N)^{a_n+1} + N]$$
$$= (-1)^{a_1+a_2+ \dots +a_n+n} N^n (N+1)^n$$
$$\times [N^{a_1} - N^{a_1-1} + N^{a_1-2} - \dots \pm 1]$$
$$\times [N^{a_2} - N^{a_2-1} + N^{a_2-2} - \dots \pm 1]$$
$$\dots\dots\dots\dots\dots\dots$$
$$[N^{a_n} - N^{a_n-1} + N^{a_n-2} - \dots \pm 1],$$

the relation (c) enables us to deduce by complete induction that the polynomial function P is divisible by $(N+1)^{n+m-1}$. Let us put

$$\mathfrak{R}\begin{bmatrix} a_1^1 & a_1^2 & \dots & a_1^n \\ a_2^1 & a_2^2 & \dots & a_2^n \\ \\ a_m^1 & a_m^2 & \dots & a_m^n \end{bmatrix} = \frac{(-1)^s}{(N+1)^{n+m-1}} \begin{bmatrix} a_1^1 & a_1^2 & \dots & a_1^n \\ a_2^1 & a_2^2 \\ \\ a_m^1 & a_m^2 & \dots & a_m^n \end{bmatrix}$$

then $\mathfrak{R}$ will be the numerator of the function connected with the multi-partition between brackets after every simplification has been made. The denominator will be in that case $(N+1)^{s-n}$.

If we put $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$, the relation (c) will become

$$\mathfrak{N}\begin{bmatrix} a_1^1 & a_1^2 & \ldots & a_1^n & \lambda_1 \\ a_2^1 & a_2^2 & \ldots & a_2^n & \lambda_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_m^1 & a_m^2 & \ldots & a_m^n & \lambda_m \end{bmatrix} = \mathfrak{N}\begin{bmatrix} \lambda_1 & \lambda_2 & \ldots & \lambda_m \\ a_1^1 & a_2^1 & \ldots & a_m^1 \\ \cdots & \cdots & \cdots & \cdots \\ a_1^n & a_2^n & \ldots & a_m^n \end{bmatrix}$$

$$= \frac{(-1)^\lambda}{N+1}[N-m+1+(-N)^{\lambda_1}+(-N)^{\lambda_2}+\ldots+(-N)^{\lambda_m}]\,\mathfrak{N}\begin{bmatrix} a_1^1 & a_1^2 & \ldots & a_1^n \\ a_2^1 & a_2^2 & \ldots & a_2^n \\ \cdots & \cdots & \cdots & \cdots \\ a_m^1 & a_m^2 & \ldots & a_m^n \end{bmatrix}$$

$$+\frac{(-1)^\lambda}{(N+1)^2}\mathbf{S}[(-N)^{\lambda_1}-1][(-N)^{\lambda_2}-1]\,\mathfrak{N}\begin{bmatrix} a_1^1+a_2^1 & a_1^2+a_2^2 & \ldots & a_1^n+a_2^n \\ a_3^1 & a_3^2 & \ldots & a_3^n \\ \cdots & \cdots & \cdots & \cdots \\ a_m^1 & a_m^2 & \ldots & a_m^n \end{bmatrix}$$

$$+\frac{(-1)^\lambda}{(N+1)^3}\mathbf{S}[(-N)^{\lambda_1}-1][(-N)^{\lambda_2}-1][(-N)^{\lambda_3}-1]$$

$$\times\,\mathfrak{N}\begin{bmatrix} a_1^1+a_2^1+a_3^1 & a_1^2+a_2^2+a_3^2 & \ldots & a_1^n+a_2^n+a_3^n \\ a_4^1 & a_4^2 & \ldots & a_4^n \\ \cdots & \cdots & \cdots & \cdots \\ a_m^1 & a_m^2 & \ldots & a_m^n \end{bmatrix} + \ldots \;\ldots\ldots\ldots\ldots(d).$$

Some particular cases of this formula will be of great use for practical purposes. The most useful are

$$\mathfrak{N}\begin{bmatrix} a_1^1 & a_1^2 & \ldots & a_1^n & a \\ a_2^1 & a_2^2 & \ldots & a_2^n & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_m^1 & a_m^2 & \ldots & a_m^n & 0 \end{bmatrix} = N(N^{a-1}-N^{a-2}+\ldots\pm 1)\,\mathfrak{N}[a] \;\ldots\ldots\ldots\ldots\ldots\ldots(\text{i}),$$

$$\mathfrak{N}\begin{bmatrix} a_1^1 & a_1^2 & \ldots & a_1^n & 1 \\ a_2^1 & a_2^2 & \ldots & a_2^n & a \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_m^1 & a_m^2 & \ldots & a_m^n & 0 \end{bmatrix} = [N^{a-1}-N^{a-2}+\ldots\pm 1][\mathfrak{N}_{12}(a)-\mathfrak{N}(a)] \;\ldots\ldots(\text{ii}),$$

$$\mathfrak{N}\begin{bmatrix} a_1^1 & a_1^2 & \ldots & a_1^n & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_p^1 & a_p^2 & \ldots & a_p^n & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_m^1 & a_m^2 & \ldots & a_m^n & 0 \end{bmatrix} = (-1)^{p-1}(p-1)\,\mathfrak{N}(a)+(-1)^p\mathbf{S}_{(p)}\,\mathfrak{N}_{12}(a)+\ldots$$

$$+(-1)^{p+1}\mathbf{S}_{(p)}\,\mathfrak{N}_{123}(a)+\ldots+\mathfrak{N}_{123\ldots p}(a)\ldots\ldots\ldots\ldots(\text{iii}).$$

The last formula becomes

$$\mathfrak{N}\begin{bmatrix} a_1^1 & a_1^2 & \ldots & a_1^n & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_p^1 & a_p^2 & \ldots & a_p^n & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_m^1 & a_m^2 & \ldots & a_m^n & 0 \end{bmatrix} = \mathfrak{N}_{123\ldots p}(a) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{iv}),$$

if $$\mathfrak{N}(a) = \mathfrak{N}_{12}(a) = \mathfrak{N}_{13}(a) = \ldots = \mathfrak{N}_{23\ldots p}(a) = 0.$$

As an example we shall calculate the coefficient of s_5^2 involved in $S(2, 4^2)$. There are three different multi-partitions to be considered:

$$\begin{array}{ccccc|c} \cdot & 1 & 1 & \cdot & \cdot & 2 \\ 1 & 1 & \cdot & 2 & \cdot & 4 \\ 1 & \cdot & 1 & \cdot & 2 & 4 \\ \hline 2 & 2 & 2 & 2 & 2 & \end{array} \qquad \begin{array}{ccc|c} (2) & (2) & & \\ \cdot & 1 & \cdot & 2 \\ 1 & \cdot & 2 & 4 \\ 1 & 1 & \cdot & 4 \\ \hline 2 & 2 & 2 & \end{array} \qquad \begin{array}{ccc|c} (3) & & & \\ \cdot & 1 & 1 & 2 \\ 1 & 1 & \cdot & 4 \\ 1 & \cdot & 1 & 4 \\ \hline 2 & 2 & 2 & \end{array}$$

The second one must be considered twice for the last rows cannot be interchanged.

The numerical coefficients are respectively

$$\frac{2!\,4!\,4!}{2!\,2!} = 288, \quad \frac{2!\,4!\,4!\,2!}{2!\,2!\,2!} = 288, \quad \frac{2!\,4!\,4!}{3!} = 192.$$

For the numerators we shall have

$$\mathfrak{N}\begin{bmatrix} \cdot & 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot & 2 \\ 1 & 1 & \cdot & 2 & \cdot \end{bmatrix} = (\mathfrak{N}(2))^2\,\mathfrak{N}\begin{bmatrix} \cdot & 1 & 1 \\ 1 & \cdot & 1 \\ 1 & 1 & \cdot \end{bmatrix} = (\mathfrak{N}(2))^3 = N^3,$$

$$\mathfrak{N}\begin{bmatrix} \cdot & \cdot & 1 & 1 & \cdot \\ 1 & 1 & \cdot & \cdot & 2 \\ 1 & 1 & 1 & 1 & \cdot \end{bmatrix} = \mathfrak{N}(2)\,\mathfrak{N}\begin{bmatrix} \cdot & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} = (\mathfrak{N}(2))^3 = N^3,$$

$$\mathfrak{N}\begin{bmatrix} \cdot & \cdot & \cdot & 1 & 1 \\ 1 & 1 & 1 & \cdot & 1 \\ 1 & 1 & 1 & 1 & \cdot \end{bmatrix} = \mathfrak{N}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = N(N^2 - N + 1);$$

hence the coefficient of s_2^5 will be

$$\frac{192\,N(4N^2 - N + 1)}{(N+1)^5}.$$

We notice that owing to the fact that the function $\mathfrak{N}$ is not altered if the columns are changed into rows and *vice versa*, we shall have the same function of N for the coefficient of $s_2 s_4^2$ involved in $S(2^5)$ (excepting perhaps the power of the denominator $N+1$). The corresponding numerical coefficient may be deduced from the previous one,

$$\frac{192}{2!\,(4!)^2\,2!} \times (2!)^5 . 5! = 320.$$

This property is quite general and is very useful for drawing up a complete table of formulae. *As a particular case we notice that the problem of working out the semi-invariants of the sampling μ's for a normal parent population is closely connected with that of the calculation of $S(2^k)$ when the sample is drawn from any population.*

As a final application of the method worked out in the present section I shall consider the general relation

$$S(2^\alpha, 3^\beta, 4^\gamma, \ldots) = K\,\frac{F(N)}{(N+1)^p}\,s_2^p,$$

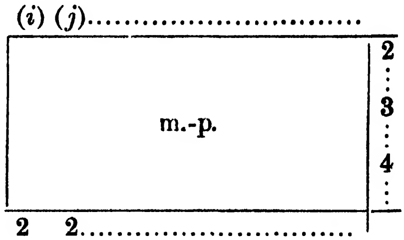
where the parent population is normal and

$$2\alpha + 3\beta + 4\gamma + \ldots = 2p.$$

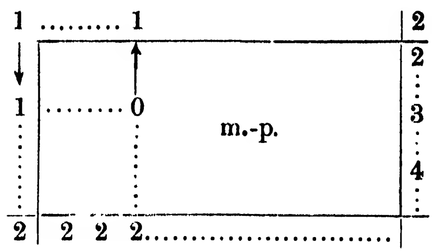
Let us consider also the relation

$$S(2^{\alpha+1}, 3^{\beta}, 4^{\gamma}, \ldots) = K_1 \frac{G(N)}{(N+1)^{p+1}} s_2^{p+1},$$

and let



be a multi-partition used for the calculation of $S(2^{\alpha}, 3^{\beta}, 4^{\gamma}, \ldots)$. We shall obtain the corresponding multi-partitions of $S(2^{\alpha+1}, 3^{\beta}, 4^{\gamma}, \ldots)$ as it is shown in the following scheme:



We notice that owing to the relation (iv) the function $\mathfrak{N}$ will be the same for both these multi-partitions; their numerical coefficients will be in the ratio $1 : 2p$. That fact enables us to write down the following general formula:

$$S(2^{\alpha}, 3^{\beta}, 4^{\gamma}, \ldots) = \left(\frac{2}{N+1}\right)^{\alpha} \frac{(p-1)!}{(p-\alpha-1)!} S(3^{\beta}, 4^{\gamma}, 5^{\delta}, \ldots) s_2^{\alpha} \quad \ldots\ldots(\text{v}),$$

which is true for samples drawn from a normal population.

Some particular cases of the formula (v) are the following:

$$S(2^{\alpha}) = \frac{2^{\alpha-1} \cdot (\alpha-1)!\, N}{(N+1)^{\alpha}} s_2^{\alpha},$$

$$S(2^{\alpha}, 4) = \frac{3 \cdot 2^{\alpha} \cdot (\alpha+1)!\, N^2}{(N+1)^{\alpha+2}} s_2^{\alpha+2},$$

$$S(2^{\alpha}, 3^2) = \frac{3 \cdot 2^{\alpha} \cdot (\alpha+2)!\, N(N-1)}{(N+1)^{\alpha+3}} s_2^{\alpha+3},$$

$$S(2^{\alpha}, 4^2) = 2^{\alpha+2} \cdot (\alpha+3)! \frac{N(4N^2 - N + 1)}{(N+1)^{\alpha+4}} s_2^{\alpha+4},$$

$$S(2^{\alpha}, 3^2, 4) = 9 \cdot 2^{\alpha} \cdot (\alpha+4)! \frac{N(N-1)(2N-1)}{(N+1)^{\alpha+5}} s_2^{\alpha+5}.$$

Normal Distributions.

14. Let us consider a normal population and take its standard deviation as unity; the corresponding characteristic function will be $e^{\frac{t^2}{2}}$. Under these hypotheses the first formula of (49) becomes

$$B_2(t_1, t_2) = \frac{1}{N}(e^{t_1 t_2} - 1) + \frac{1}{N^2}(e^{t_1^2 + t_1 t_2} + e^{t_2^2 + t_1 t_2} - \tfrac{1}{2} e^{t_1^2 + 2t_1 t_2} - \tfrac{1}{2} e^{t_2^2 + 2t_1 t_2}$$
$$+ \tfrac{1}{2} e^{2t_1 t_2} - e^{t_1 t_2} - \tfrac{1}{2} e^{t_1^2} - \tfrac{1}{2} e^{t_2^2} + \tfrac{1}{2}) + \dots \qquad \text{(69)}.$$

We have, on the other hand,

$$\left.\begin{aligned} e^{t_1^2 + t_1 t_2} &= \sum_{0}^{\infty}{}_p \sum_{0}^{p}{}_k \frac{(p+k)!}{k!} \cdot \frac{t_1^{p+k}}{(p+k)!} \cdot \frac{t_2^{p-k}}{(p-k)!} \\ e^{t_1^2 + 2t_1 t_2} &= \sum_{0}^{\infty}{}_p \sum_{0}^{p}{}_k 2^{p-k} \frac{(p+k)!}{k!} \cdot \frac{t_1^{p+k}}{(p+k)!} \cdot \frac{t_2^{p-k}}{(p-k)!} \end{aligned}\right\} \qquad \text{(70)},$$

where we must distinguish three possibilities:

(1) $a + b = 2p + 1$. In that case

$$L_{a,b} = 0 \qquad \text{(71)}.$$

(2) $a + b = 2p$, $a - b = 2k$, $k > 0$. Owing to the relations (70) we obtain

$$L_{a,b} = -\frac{p+k!}{k!} \cdot \frac{2^{p-k-1} - 1}{N^2} + \dots {}^{*} \qquad \text{(72)}.$$

(3) $a = b = p$. In the same manner, as before, we find

$$L_{a,b} = \frac{p!}{N}\left(1 - \frac{2^p - 2^{p-1} - 1}{N} + \dots\right) \qquad \text{(73)}.$$

For the associated function of the third order semi-invariant we have

$$B_3(t_1, t_2, t_3) = \frac{1}{N^2}(e^{t_1 t_2 + t_2 t_3 + t_3 t_1} - e^{t_1(t_2 + t_3)} - e^{t_2(t_3 + t_1)}$$
$$- e^{t_3(t_1 + t_2)} + e^{t_1 t_2} + e^{t_2 t_3} + e^{t_3 t_1} - 1) + \dots \qquad \text{(74)},$$

where we distinguish the possibilities:

(1) $a = \beta + \gamma$, $b = \gamma + \alpha$, $c = \alpha + \beta$; α, β, γ, being different from 0, in which case

$$L_{a,b,c} = \frac{(\alpha+\beta)!}{\gamma!} \cdot \frac{(\beta+\gamma)!}{\alpha!} \cdot \frac{(\gamma+\alpha)!}{\beta!} \cdot \frac{1}{N^2} + \dots \qquad \text{(75)}.$$

(2) In all the other cases

$$L_{a,b,c} = O\left(\frac{1}{N^2}\right) \qquad \text{(76)}.$$

Finally we find

$$L_{a,b,c,d} = \frac{\alpha!\,\beta!}{N^2} + \dots \qquad \text{(77)},$$

if the indices a, b, c, d are equal two by two, and $L_{a,b,c,d} = O\left(\frac{1}{N^2}\right)$ in any other case.

* Again, if $b = 1$, $L_{a,b} = O\left(\frac{1}{N^2}\right)$.

15. More complete results can be obtained for the distribution of the sampling moments about the mean. In order to work out the corresponding formulae, I shall consider first the development of the expression

$$e^{s(t_1^2+t_2^2)+r t_1 t_2} = \sum_0^\infty \mathbf{S} \frac{s^{a+c} r^b}{a!\,b!\,c!} t_1^{2a+b} t_2^{2c+b} \quad \text{.................(78)},$$

since the coefficient of $\frac{t_1^m}{m!} \cdot \frac{t_2^m}{n!}$ is found to be

$$m!\,n!\,\mathbf{S} \frac{s^{a+c} r^b}{a!\,b!\,c!} \quad \text{..............................(79)},$$

the sum being extended over all the non-negative integers for which $m = 2a + b$, $n = 2c + b$. Therefore the coefficient considered is zero, unless m and n are both at the same time even or odd, i.e. $|m - n| = 2d$. Obviously, owing to the symmetry of the expression (78), we can suppose $m \geqslant n$; the sum (79) may be written down as follows:

$$m!\,n!\,s^d r^n \sum_0^p{}_q \frac{1}{q!\,(d+q)!\,(n-2q)!} \left(\frac{s}{r}\right)^{2q} \quad \text{.................(80)},$$

where p stands for the greatest integer contained in n*.

Since we have

$$(N-2)\,\phi\left(-\frac{t_1+t_2}{N}\right) + \phi\left(t_1 - \frac{t_1+t_2}{N}\right) + \phi\left(t_2 - \frac{t_1+t_2}{N}\right) = \frac{N-1}{2N}(t_1^2+t_2^2) - \frac{1}{N} t_1 t_2,$$

$$(N-1)\,\phi\left(-\frac{t_1+t_2}{N}\right) + \phi\left(t_1 + t_2 - \frac{t_1+t_2}{N}\right) = \frac{N-1}{2N}(t_1^2+t_2^2+2t_1t_2),$$

by applying the previous result to the development on the right-hand side of (63), we obtain

$$S(m,n) = \frac{m!\,n!}{2^d . N}\left(1 - \frac{1}{N}\right)^{d+1}$$

$$\times \sum_0^{p-1}{}_q \frac{1}{q!\,(d+q)!\,(n-2q)!} \cdot \frac{1}{2^{2c}} \left[(N-1)^{2c}\left(-\frac{1}{N}\right)^{n-1} + \left(1 - \frac{1}{N}\right)^{n-1}\right] \text{...(81)},$$

for when there is a term following from the last term on the right-hand side of (63), this term is reduced by the terms corresponding to $q = p$; when such a term does not exist, the latter terms cancel out.

16. The formulae obtained above (§ 14) show us that the semi-invariants of a sample from a normal population are independent, as far as the first power of $\frac{1}{N}$.

As a final application of the method of associated functions, I propose to show now that this is a specific property of the samples from a normal population.

In order that the sampling semi-invariants be independent, owing to the first formula (48), we have the condition

$$\phi(t_1 + t_2) - \phi(t_1) - \phi(t_2) = F(t_1 t_2) \quad \text{...................(82)};$$

hence

$$\phi(2t) - \phi(-2t) = 2\,[\phi(t) - \phi(-t)].$$

* This result gives, at the same time, the expression of the product moments of a normal distribution of two variates, in terms of the corresponding correlation coefficient.

This is a well-known functional equation, and its solution is

$$\phi(t) - \phi(-t) = s_1 t,$$

which gives

$$\phi(t) = s_1 t + \psi(t^2).$$

Substituting this function in the relation (82), we obtain

$$\psi[(t_1 + t_2)^2] - \psi(t_1^2) - \psi(t_2^2) = F(t_1 t_2).$$

If we differentiate this relation with respect to t_1 and make afterwards $t_1 = 0$, we obtain

$$2\psi'(t_2^2) = F'(0);$$

hence $\psi(t) = s_2 \dfrac{t}{2}$ and consequently $\phi(t) = s_1 \dfrac{t}{1!} + s_2 \dfrac{t^2}{2!}$.

Conclusions.

In this paper I have introduced new functions connected with the distributions of random variables and indicated a new procedure, based upon them, for obtaining further results in the sampling problem, although the same method may be applied for calculating in an easier way formulae already worked out by previous authors.

I have shown that having reached the general expression of the associated functions we are able to write down any desired formula regarding the distribution of sample characteristics. However, a certain limit must be considered, beyond which the calculations become too complicated and the labour required to obtain some of these formulae seems out of proportion to the practical importance of any applications which can be made of them. The general formulae present nevertheless an advantage which has not been reached before—at least as far as I know—and this advantage is double. First they enable us to calculate any formula with any degree of approximation, and this, by a methodical procedure. On the other hand, they may be very useful for theoretical researches, for they embrace in one entirety all the moments—or semi-invariants—of the same order.

Besides, by using the associated functions, I have elucidated one point regarding the sampling semi-invariants showing that generally they are not independent—as one might think; the only case in which they are independent is that of very large samples from a normal population.

Note.

Let us consider n random variables $x_1, x_2, \dots x_n$ and suppose that between them there is the following relation:

$$x_1 + x_2 + \dots + x_n \quad a.$$

In that case no real law of distribution can exist for the ensemble of these variables unless we consider $n - 1$ of them only. However, there is a characteristic function for the whole system of variables, which has all the properties of a general function such as this, except that which allows us to calculate by an integration the corresponding law of distribution.

Let $F(x_1, x_2, \dots x_{n-1})\, dx_1\, dx_2 \dots dx_{n-1}$ be the distribution of $x_1, x_2, \dots x_{n-1}$ and $\phi(t_1, t_2, \dots t_{n-1})$ the characteristic function

$$\phi(t_1, t_2, \dots t_{n-1}) = \int_{(D)} F(x_1, x_2, \dots x_{n-1})\, e^{t_1 x_1 + t_2 x_2 + \dots + t_{n-1} x_{n-1}}\, dx_1\, dx_2 \dots dx_{n-1}.$$

Now, if we make abstraction for a moment of the relation between the x's, we have

$$f(u_1, u_2, \dots u_n) = \int_{(D)} F e^{u_1 x_1 + u_2 x_2 + \dots + u_n x_n}\, dx_1\, dx_2 \dots dx_{n-1},$$

and putting $a - x_1 - x_2 - \dots - x_{n-1}$ instead of x_n, we shall find for the characteristic function of $x_1, x_2, \dots x_n$:

$$f(u_1, u_2, \dots u_n) = e^{a}\, \phi(u_1 - u_n, u_2 - u_n, \dots u_{n-1} - u_n) \quad \dots\dots\dots\dots (1).$$

Let us apply this result to the distribution of $\delta\varpi_1, \delta\varpi_2, \dots \delta\varpi_n$. We know that the characteristic function of $\delta\varpi_1, \delta\varpi_2, \dots \delta\varpi_{n-1}$ is*

$$\phi(t_1, t_2, \dots t_{n-1}) = e^{-\Sigma p_i t_i}\,[p_n + \Sigma p_i e^{\frac{t_i}{N}}]^N,$$

and since between the $\delta\varpi_i$'s we have the relation

$$\delta\varpi_1 + \delta\varpi_2 + \dots + \delta\varpi_n = 0,$$

by using the formula (1) we find for the characteristic function of $\delta\varpi_1, \delta\varpi_2, \dots \delta\varpi_n$ the result given before (§ 2, (1)), namely:

$$f(u_1, u_2, \dots u_n) = e^{-\Sigma p_i u_i}\,[\Sigma p_i e^{\frac{u_i}{N}}]^N,$$

where n replaces the t of the formula there given.

APPENDIX I.

Table of Formulae giving $L^{(2)}_{a,b}$†, the Second Order Semi-invariants of the Sampling Semi-invariants. Size of Sample $= N$.

$$L_{1,1} = \frac{1}{N} s_2,$$

$$L_{1,2} = \left(\frac{1}{N} - \frac{1}{N^2}\right) s_3,$$

$$L^{(2)}_{1,3} = \left(\frac{1}{N} - \frac{3}{N^2}\right) s_4,$$

$$L^{(2)}_{1,4} = \left(\frac{1}{N} - \frac{7}{N^2}\right) s_5 - \frac{12}{N^2} s_2 s_3,$$

$$L^{(2)}_{1,5} = \left(\frac{1}{N} - \frac{15}{N^2}\right) s_6 - \frac{60}{N^2} s_2 s_4 - \frac{60}{N^2} s_3^2,$$

$$L^{(2)}_{1,6} = \left(\frac{1}{N} - \frac{31}{N^2}\right) s_7 - \frac{210}{N^2} s_2 s_5 - \frac{570}{N^2} s_3 s_4 - \frac{960}{N^2} s_2^2 s_3.$$

* See G. Darmois, *Statistique mathématique* (Paris, G. Doin, 1928), pp. 237—239.

† $L^2(a, b)$ denotes approximate values of $L(a, b)$ as far as terms in $1/N^2$.

$$L^{(2)}_{2,2}=\left(\frac{1}{N}-\frac{2}{N^2}\right)s_4+2\left(\frac{1}{N}-\frac{1}{N^2}\right)s_2^2,$$

$$L^{(2)}_{2,3}=\left(\frac{1}{N}-\frac{4}{N^2}\right)s_5+6\left(\frac{1}{N}-\frac{3}{N^2}\right)s_2s_3,$$

$$L^{(2)}_{2,4}=\left(\frac{1}{N}-\frac{8}{N^2}\right)s_6+\left(\frac{1}{N}-\frac{7}{N^2}\right)(8s_2s_4+6s_3^2)-\frac{12}{N^2}s_2s_4-\frac{24}{N^2}s_2^3,$$

$$L^{(2)}_{2,5}=\left(\frac{1}{N}-\frac{16}{N^2}\right)s_7+\left(\frac{1}{N}-\frac{15}{N^2}\right)(10s_2s_5+20s_3s_4)-\frac{60}{N^2}s_2s_5-\frac{60}{N^2}s_3s_4-\frac{480}{N^2}s_2^2s_3,$$

$$L^{(2)}_{2,6}=\left(\frac{1}{N}-\frac{32}{N^2}\right)s_8+\left(\frac{1}{N}-\frac{31}{N^2}\right)(12s_2s_6+30s_3s_5+20s_4^2)-\frac{210}{N^2}s_2s_6-\frac{360}{N^2}s_3s_5$$
$$-\frac{210}{N^2}s_4^2-\frac{2280}{N^2}s_2^2s_4-\frac{3420}{N^2}s_2s_3^2-\frac{360}{N^2}s_2^4.$$

$$L^{(2)}_{3,3}=\left(\frac{1}{N}-\frac{6}{N^2}\right)s_6+\left(\frac{1}{N}-\frac{5}{N^2}\right)(9s_2s_4+9s_3^2)+6\left(\frac{1}{N}-\frac{3}{N^2}\right)s_2^3,$$

$$L^{(2)}_{3,4}=\left(\frac{1}{N}-\frac{10}{N^2}\right)s_7+\left(\frac{1}{N}-\frac{9}{N^2}\right)(12s_2s_5+30s_3s_4)+36\left(\frac{1}{N}-\frac{7}{N^2}\right)s_2^2s_3-\frac{12}{N^2}s_2s_5-\frac{72}{N^2}s_2^2s_3,$$

$$L^{(2)}_{3,5}=\left(\frac{1}{N}-\frac{18}{N^2}\right)s_8+\left(\frac{1}{N}-\frac{17}{N^2}\right)(15s_2s_6+45s_3s_5+30s_4^2)+\left(\frac{1}{N}-\frac{15}{N^2}\right)(60s_2^2s_4+90s_2s_3^2)$$
$$\frac{60}{N^2}s_2s_6-\frac{60}{N^2}s_3s_5-\frac{540}{N^2}s_2^2s_4-\frac{900}{N^2}s_2s_3^2-\frac{360}{N^2},$$

$$L^{(2)}_{3,6}=\left(\frac{1}{N}-\frac{34}{N^2}\right)s_9+\left(\frac{1}{N}-\frac{33}{N^2}\right)(18s_2s_7+63s_3s_6+105s_4s_5)$$
$$+\left(\frac{1}{N}-\frac{31}{N^2}\right)(90s_2^2s_5+360s_2s_3s_4+90s_3^2)-\frac{210}{N^2}s_2s_7-\frac{360}{N^2}s_3s_6$$
$$-\frac{210}{N^2}s_4s_5-\frac{2520}{N^2}s_2^2s_5-\frac{10800}{N^2}s_2s_3s_4-\frac{3240}{N^2}s_3^2-\frac{9720}{N^2}s_2^3s_3.$$

$$L^{(2)}_{4,4}=\left(\frac{1}{N}-\frac{14}{N^2}\right)s_8+\left(\frac{1}{N}-\frac{13}{N^2}\right)(16s_2s_6+48s_3s_5+34s_4^2)+\left(\frac{1}{N}-\frac{11}{N^2}\right)(72s_2^2s_4+144s_2s_3^2)$$
$$+24\left(\frac{1}{N}-\frac{7}{N^2}\right)s_2^4-\frac{24}{N^2}s_2s_6-\frac{192}{N^2}s_2^2s_4-\frac{144}{N^2}s_2s_3^2,$$

$$L^{(2)}_{4,5}=\left(\frac{1}{N}-\frac{22}{N^2}\right)s_9+\left(\frac{1}{N}-\frac{21}{N^2}\right)(20s_2s_7+70s_3s_6+120s_4s_5)$$
$$+\left(\frac{1}{N}-\frac{19}{N^2}\right)(120s_2^2s_5+600s_2s_3s_4+180s_3^3)+240\left(\frac{1}{N}-\frac{15}{N^2}\right)s_2^3s_3$$
$$-\frac{72}{N^2}s_2s_7-\frac{60}{N^2}s_3s_6-\frac{840}{N^2}s_2^2s_5-\frac{2520}{N^2}s_2s_3s_4-\frac{360}{N^2}s_3^3-\frac{4340}{N^2}s_2^3s_3,$$

$$L^{(2)}_{4,6}=\left(\frac{1}{N}-\frac{38}{N^2}\right)s_{10}+\left(\frac{1}{N}-\frac{37}{N^2}\right)(24s_2s_8+96s_3s_7+194s_4s_6+120s_5^2)$$
$$+\left(\frac{1}{N}-\frac{35}{N^2}\right)(180s_2^2s_6+1080s_2s_3s_5+720s_2s_4^2+1260s_3^2s_4)$$
$$+\left(\frac{1}{N}-\frac{31}{N^2}\right)(480s_2^3s_4+1080s_2^2s_3^2)-\frac{222}{N^2}s_2s_8-\frac{360}{N^2}s_3s_7-\frac{3504}{N^2}s_2^2s_6$$
$$-\frac{11880}{N^2}s_2s_3s_5-\frac{9060}{N^2}s_2s_4^2-\frac{12060}{N^2}s_3^2s_4-\frac{17640}{N^2}s_2^3s_4-\frac{43200}{N^2}s_2^2s_3^2-\frac{5040}{N^2}s_2^5.$$

$$
\begin{aligned}
L^{(2)}_{5,5} = {} & \left(\frac{1}{N} - \frac{30}{N^2}\right) s_{10} + \left(\frac{1}{N} - \frac{29}{N^2}\right)(25 s_2 s_8 + 100 s_3 s_7 + 200 s_4 s_6 + 125 s_5^2) \\
& + \left(\frac{1}{N} - \frac{27}{N^2}\right)(1200 s_2 s_3 s_5 + 200 s_2^3 s_6 + 850 s_2 s_4^2 + 1500 s_3^2 s_4) \\
& + \left(\frac{1}{N} - \frac{23}{N}\right)(1800 s_2^2 s_3^2 + 600 s_2^3 s_4) + 120 \left(\frac{1}{N} - \frac{15}{N^3}\right) s_2^5 - \frac{120}{N^2} s_2 s_8 - \frac{120}{N^2} s_3 s_7 \\
& - \frac{6600}{N^2} s_2 s_3 s_5 - \frac{1800}{N^2} s_2^2 s_6 - \frac{3600}{N^2} s_2 s_4^2 - \frac{2400}{N^2} s_3^2 s_4 - \frac{10800}{N^2} s_2^2 s_3^2 - \frac{7200}{N^2} s_2^3 s_4,
\end{aligned}
$$

$$
\begin{aligned}
L^{(2)}_{5,6} = {} & \left(\frac{1}{N} - \frac{46}{N^2}\right) s_{11} + \left(\frac{1}{N} - \frac{45}{N^2}\right)(30 s_2 s_9 + 135 s_3 s_8 + 310 s_4 s_7 + 455 s_5 s_6) \\
& + \left(\frac{1}{N} - \frac{43}{N^2}\right)(2100 s_2 s_3 s_6 + 3600 s_2 s_4 s_5 + 300 s_2^2 s_7 + 3150 s_3^2 s_5 + 3150 s_4^2 s_3) \\
& + \left(\frac{1}{N} - \frac{39}{N^2}\right)(9000 s_2^2 s_3 s_4 + 5400 s_3^3 s_2 + 1200 s_2^3 s_5) + 1800 \left(\frac{1}{N} - \frac{31}{N^2}\right) s_2^4 s_3 \\
& - \frac{270}{N^2} s_2 s_9 - \frac{420}{N^2} s_3 s_8 - \frac{24600}{N^2} s_2 s_3 s_6 - \frac{33600}{N^2} s_2 s_4 s_5 - \frac{5640}{N^2} s_2^2 s_7 \\
& - \frac{18000}{N^2} s_3^2 s_5 - \frac{151200}{N^2} s_2^2 s_3 s_4 - \frac{12420}{N^2} s_3^3 s_2 - \frac{30600}{N^2} s_2^3 s_5 - \frac{50400}{N^2} s_2^4 s_3.
\end{aligned}
$$

$$
\begin{aligned}
L^{(2)}_{6,6} = {} & \left(\frac{1}{N} - \frac{62}{N^2}\right) s_{12} + \left(\frac{1}{N} - \frac{61}{N^2}\right)(36 s_2 s_{10} + 180 s_3 s_9 + 465 s_4 s_8 + 780 s_5 s_7 + 461 s_6^2) \\
& + \left(\frac{1}{N} - \frac{59}{N^2}\right)(450 s_2^2 s_8 + 6300 s_3^2 s_6 + 4950 s_4^3 + 3600 s_2 s_3 s_7 \\
& \qquad + 7200 s_2 s_4 s_6 + 4500 s_2 s_5^2 + 21600 s_3 s_4 s_5) \\
& + \left(\frac{1}{N} - \frac{55}{N^2}\right)(21600 s_2^2 s_3 s_5 + 15300 s_2^2 s_4^2 + 2400 s_2^3 s_6 + 8100 s_3^4 + 54000 s_2 s_3^2 s_4) \\
& + \left(\frac{1}{N} - \frac{47}{N^2}\right)(21600 s_2^3 s_3^2 + 5400 s_2^4 s_4) + 720 \left(\frac{1}{N} - \frac{31}{N^2}\right) s_2^6 - \frac{420}{N^2} s_2 s_{10} - \frac{480}{N^2} s_3 s_9 \\
& - \frac{420}{N^2} s_4 s_8 - \frac{10440}{N^2} s_2^2 s_8 - \frac{45360}{N^2} s_3^2 s_6 - \frac{8400}{N^2} s_4^3 - \frac{49880}{N^2} s_2 s_3 s_7 - \frac{85240}{N^2} s_2 s_4 s_6 \\
& - \frac{41600}{N^2} s_2 s_5^2 - \frac{88200}{N^2} s_3 s_4 s_5 - \frac{540000}{N^2} s_2^2 s_3 s_5 - \frac{309600}{N^2} s_2^2 s_4^2 - \frac{75600}{N^2} s_2^3 s_6 \\
& - \frac{74800}{N^2} s_3^4 - \frac{788400}{N^2} s_2 s_3^2 s_4 - \frac{453600}{N^2} s_2^3 s_3^2 - \frac{201600}{N^2} s_2^4 s_4.
\end{aligned}
$$

APPENDIX II.

Table of Semi-invariants of the Sampling Product Moments about their Mean.

Tables giving $S(2^\alpha, 3^\beta, 4^\gamma)$ obtained so far.

N.B. The solidus in the expressions below does not extend to the s powers and products. Further for greater brevity of expression the size of the sample is taken to be $(N + 1)$.

Weight 2

$S(2) = N/(N+1)\, s_2$*.

Weight 3

$S(3) = N(N-1)/(N+1)^2\, s_3$*.

* Formulae marked by an asterisk have certainly been given before. This statement does not of course cover R. A. Fisher's Semi-invariants of the sampling cumulative moments. He also proceeds in that case to weight 12.

Weight 4

$S(4) = N(N^2-N+1)/(N+1)^3 s_4 + 3N^2/(N+1)^2 s_2^2$*.

$S(2^2) = N^2/(N+1)^3 s_4 + 2N/(N+1)^2 s_2^2$*.

Weight 5

$S(2,3) = N^2(N-1)/(N+1)^4 s_5 + 6N(N-1)/(N+1)^3 s_2 s_3$*.

Weight 6

$S(2,4) = N^2(N^2-N+1)/(N+1)^5 s_6 + 2N(7N^2-4N+4)/(N+1)^4 s_2 s_4$
$\quad + 6N(N-1)^2/(N+1)^4 s_3^2 + 12N^2/(N+1)^3 s_2^3$*.

$S(3^2) = N^2(N-1)^2/(N+1)^5 s_6 + 9N(N-1)^2/(N+1)^4 s_2 s_4$
$\quad + 9N(N-1)^2/(N+1)^4 s_3^2 + 6N(N-1)/(N+1)^3 s_2^3$*.

$S(2^3) = N^3/(N+1)^5 s_6 + 12N^2/(N+1)^4 s_2 s_4 + 4N(N-1)/(N+1)^4 s_3^2 + 8N/(N+1)^3 s_2^3$*.

Weight 7

$S(3,4) = N^2(N-1)(N^2-N+1)/(N+1)^6 s_7 + 6N(N-1)(3N^2-2N+2)/(N+1)^5 s_2 s_5$
$\quad + 6N(N-1)(5N^2-8N+5)/(N+1)^5 s_3 s_4 + 36N(N-1)(2N-1)/(N+1)^4 s_2^2 s_3$*.

$S(2^2,3) = N^3(N-1)/(N+1)^6 s_7 + 16N^2(N-1)/(N+1)^5 s_2 s_5$
$\quad + 12N(N-1)(2N-1)/(N+1)^5 s_3 s_4 + 48N(N-1)/(N+1)^4 s_2^2 s_3$*.

Weight 8

$S(4^2) = N^2(N^2-N+1)^2/(N+1)^7 s_8 + 4N(N^2-N+1)(7N^2-4N+4)/(N+1)^6 s_2 s_6$
$\quad + 48N(N-1)^3(N^2-N+1)/(N+1)^6 s_3 s_5 + 2N(17N^4-43N^3+78N^2-52N+17)/(N+1)^6 s_4^2$
$\quad + 12N(17N^3-20N^2+26N-6)/(N+1)^5 s_2^2 s_4 + 72N(N-1)^2(3N-2)/(N+1)^5 s_2 s_3^2$
$\quad + 24N(4N^2-N+1)/(N+1)^4 s_2^4$*.

$S(2^2,4) = N^3(N^2-N+1)/(N+1)^7 s_8 + 2N^2(13N^2-10N+10)/(N+1)^6 s_2 s_6$
$\quad + 8N(N-1)(5N^2-5N+2)/(N+1)^6 s_3 s_5 + 2N(17N^3-20N^2+26N-6)/(N+1)^6 s_4^2$
$\quad + 16N(11N^2-5N+5)/(N+1)^5 s_2^2 s_4 + 24N(N-1)(6N-5)/(N+1)^5 s_2 s_3^2 + 72N^2/(N+1)^4 s_2^4$*.

$S(2,3^2) = N^3(N-1)^2/(N+1)^7 s_8 + 21N^2(N-1)^2/(N+1)^6 s_2 s_6 + 6N(N-1)^2(8N-3)/(N+1)^6 s_3 s_5$
$\quad + 9N(N-1)^2(3N-2)/(N+1)^6 s_4^2 + 18N(N-1)(6N-5)/(N+1)^5 s_2^2 s_4$
$\quad + 18N(N-1)(9N-11)/(N+1)^5 s_2 s_3^2 + 36N(N-1)/(N+1)^4 s_2^4$.

$S(2^4) = N^4/(N+1)^7 s_8 + 24N^3/(N+1)^6 s_2 s_6 + 32N^2(N-1)/(N+1)^6 s_3 s_5$
$\quad + 8N(4N^2-N+1)/(N+1)^6 s_4^2 + 144N^2/(N+1)^5 s_2^2 s_4 + 96N(N-1)/(N+1)^5 s_2 s_3^2$
$\quad + 48N/(N+1)^4 s_2^4$*.

Weight 9

$S(2,3,4) = N^3(N-1)(N^2-N+1)/(N+1)^8 s_9 + 2N^2(N-1)(16N^2-13N+13)/(N+1)^7 s_2 s_7$
$\quad + 6N(N-1)(12N^3-21N^2+16N-4)/(N+1)^7 s_3 s_6$
$\quad + 2N(N-1)(58N^3-91N^2+103N-30)/(N+1)^7 s_4 s_5$
$\quad + 12N(N-1)(16N^2-19N+16)/(N+1)^6 s_2^2 s_5$
$\quad + 12N(N-1)(64N^2-94N+67)/(N+1)^6 s_2 s_3 s_4 + 36N(N-1)^2(5N-7)/(N+1)^6 s_3^3$
$\quad + 360N(N-1)(2N-1)/(N+1)^5 s_2^3 s_3$.

$S(3^3) = N^3(N-1)^3/(N+1)^8 s_9 + 27N^2(N-1)^3/(N+1)^7 s_2 s_9 + 27N(N-1)^2(3N-1)/(N+1)^7 s_3 s_6$
$\quad + 27N(N-1)^2(4N-3)/(N+1)^7 s_4 s_5 + 54N(N-1)^2(4N-3)/(N+1)^6 s_2^2 s_5$
$\quad + 162N(N-1)(5N-7)/(N+1)^6 s_2 s_3 s_4 + 36N(N-1)(7N^2-16N+11)/(N+1)^6 s_3^3$
$\quad + 108N(N-1)(5N-7)/(N+1)^5 s_2^3 s_3$*.

$S(2^3,3) = N^4(N-1)/(N+1)^8 s_9 + 30N^3(N-1)/(N+1)^7 s_2 s_7 + 2N^2(N-1)(31N-22)/(N+1)^7 s_3 s_6$
$\quad + 12N(N-1)(9N^2-5N+2)/(N+1)^7 s_4 s_5 + 240N^2(N-1)/(N+1)^6 s_2^2 s_5$
$\quad + 360N(N-1)(2N-1)/(N+1)^6 s_2 s_3 s_4 + 24N(N-1)(5N-7)/(N+1)^6 s_3^3$
$\quad + 480N(N-1)/(N+1)^5 s_2^3 s_3$.

* See footnote on p. 103.

Weight 10

$$S(2,4^2) = N^3(N^2-N+1)^2/(N+1)^9 s_{10} + 4N^2(N^2-N+1)(11N^2-8N+8)/(N+1)^8 s_2 s_8$$
$$+ 8N(N-1)(N^2-N+1)(7N^2-7N+4)/(N+1)^8 s_3 s_7$$
$$+ 4N(52N^5-134N^4+249N^3-212N^2+121N-24)/(N+1)^8 s_4 s_6$$
$$+ 4N(N-1)(29N^4-58N^3+84N^2-64N+17)/(N+1)^8 s_5^2$$
$$+ 4N(149N^4-214N^3+288N^2-130N+56)/(N+1)^7 s_2^2 s_6$$
$$+ 96N(N-1)(21N^3-37N^2+34N-16)/(N+1)^7 s_2 s_3 s_5$$
$$+ 72N(21N^4-44N^3+76N^2-48N+16)/(N+1)^7 s_2 s_4^2$$
$$+ 72N(N-1)(22N^3-58N^2+67N-29)/(N+1)^7 s_3^2 s_4$$
$$+ 48N(59N^3-62N^2+80N-18)/(N+1)^6 s_2^3 s_4$$
$$+ 144N(N-1)(29N^2-45N+19)/(N+1)^6 s_2^2 s_3^2 + 192N(4N^2-N+1)/(N+1)^5 s_2^5.$$

$$S(3^2,4) = N^3(N-1)^2(N^2-N+1)/(N+1)^9 s_{10} + 3N^2(N-1)^2(13N^2-11N+11)/(N+1)^8 s_2 s_8$$
$$+ 6N(N-1)^2(19N^3-31N^2+25N-6)/(N+1)^8 s_3 s_7$$
$$+ 3N(N-1)^2(65N^3-119N^2+125N-42)/(N+1)^8 s_4 s_6$$
$$+ 6N(N-1)^2(20N^3-34N^2+43N-15)/(N+1)^8 s_5^2$$
$$+ 18N(N-1)(26N^3-44N^2+33N-14)/(N+1)^7 s_2^2 s_6$$
$$+ 36N(N-1)(56N^3-133N^2+131N-58)/(N+1)^7 s_2 s_3 s_5$$
$$+ 54N(N-1)(22N^3-58N^2+67N-29)/(N+1)^7 s_2 s_4^2$$
$$+ 54N(N-1)(33N^3-105N^2+125N-57)/(N+1)^7 s_3^2 s_4$$
$$+ 36N(N-1)(64N^2-91N+46)/(N+1)^6 s_2^3 s_4$$
$$+ 108N(N-1)(38N^2-79N+43)/(N+1)^6 s_2^2 s_3^2 + 216N(N-1)(2N-1)/(N+1)^5 s_2^5.$$

$$S(2^3,4) = N^4(N^2-N+1)/(N+1)^9 s_{10} + 6N^3(7N^2-6N+6)/(N+1)^8 s_2 s_8$$
$$+ 4N^2(N-1)(23N^2-23N+14)/(N+1)^8 s_3 s_7$$
$$+ 2N(103N^4-146N^3+199N^2-68N+16)/(N+1)^8 s_4 s_6$$
$$+ 12N(N-1)(9N^3-9N^2+8N-2)/(N+1)^8 s_5^2 + 60N^2(9N^2-6N+6)/(N+1)^7 s_2^2 s_6$$
$$+ 96N(N-1)(17N^2-15N+6)/(N+1)^7 s_2 s_3 s_5 + 24N(59N^3-62N^2+80N-18)/(N+1)^7 s_2 s_4^2$$
$$+ 24N(N-1)(64N^2-91N+46)/(N+1)^7 s_3^2 s_4 + 480N(5N^2-2N+2)/(N+1)^6 s_2^3 s_4$$
$$+ 720N(N-1)(4N-3)/(N+1)^6 s_2^2 s_3^2 + 576N^2/(N+1)^5 s_2^5.$$

$$S(2^2,3^2) = N^4(N-1)^2/(N+1)^9 s_{10} + 37N^3(N-1)^2/(N+1)^8 s_2 s_8$$
$$+ 6N^2(N-1)^2(17N-10)/(N+1)^8 s_3 s_7 + 3N(N-1)^2(61N^2-44N+12)/(N+1)^8 s_4 s_6$$
$$+ 2N(N-1)^2(59N^2-36N+18)/(N+1)^8 s_5^2 + 6N^2(N-1)(67N-64)/(N+1)^7 s_2^2 s_6$$
$$+ 24N(N-1)(71N^2-104N+27)/(N+1)^7 s_2 s_3 s_5$$
$$+ 36N(N-1)(29N^2-45N+19)/(N+1)^7 s_2 s_4^2$$
$$+ 36N(N-1)(38N^2-79N+43)/(N+1)^7 s_3^2 s_4 + 360N(N-1)(4N-3)/(N+1)^6 s_2^3 s_4$$
$$+ 216N(N-1)(13N-17)/(N+1)^6 s_2^2 s_3^2 + 288N(N-1)/(N+1)^5 s_2^5.$$

$$S(2^5) = N^5/(N+1)^9 s_{10} + 40N^4/(N+1)^8 s_2 s_8 + 80N^3(N-1)/(N+1)^8 s_3 s_7$$
$$+ 40N^2(5N^2-2N+2)/(N+1)^8 s_4 s_6 + 16N(N-1)(6N^2+1)/(N+1)^8 s_5^2$$
$$+ 180N^3/(N+1)^7 s_2^2 s_6 + 1280N^2(N-1)/(N+1)^7 s_2 s_3 s_5 + 320N(4N^2-N+1)/(N+1)^7 s_2 s_4^2$$
$$+ 480N(N-1)(2N-1)/(N+1)^7 s_3^2 s_4 + 1920N^2/(N+1)^6 s_2^3 s_4$$
$$+ 1920N(N-1)/(N+1)^6 s_2^2 s_3^2 + 384N/(N+1)^5 s_2^5.$$

Weight 11

$$S(3,4^2) = N^3(N-1)(N^2-N+1)^2/(N+1)^{10} s_{11}$$
$$+ 4N^2(N-1)(N^2-N+1)(13N^2-10N+10)/(N+1)^9 s_2 s_9$$
$$+ 12N(N-1)(N^2-N+1)(13N^3-24N^2+17N-4)/(N+1)^9 s_3 s_8$$
$$+ 4N(N-1)(80N^5-226N^4+408N^3-367N^2+248N-48)/(N+1)^9 s_4 s_7$$
$$+ 4N(N-1)(112N^5-320N^4+579N^3-608N^2+373N-87)/(N+1)^9 s_5 s_6$$
$$+ 12N(N-1)(73N^4-114N^3+152N^2-76N+32)/(N+1)^8 s_2^2 s_7$$
$$+ 24N(N-1)(169N^4-434N^3+573N^2-410N+154)/(N+1)^8 s_2 s_3 s_6$$
$$+ 24N(N-1)(258N^4-652N^3+1079N^2-836N+304)/(N+1)^8 s_2 s_4 s_5$$

Weight 11 (*contd.*)

$$
\begin{aligned}
&+72N(N-1)^2(56N^3-146N^2+167N-97)/(N+1)^8 s_3{}^2 s_5\\
&+24N(N-1)(215N^4-750N^3+1290N^2-1152N+431)/(N+1)^8 s_3 s_4{}^2\\
&+48N(N-1)(116N^3-155N^2+182N-72)/(N+1)^7 s_2{}^3 s_5\\
&+144N(N-1)(190N^3-429N^2+495N-200)/(N+1)^7 s_2{}^2 s_3 s_4\\
&+144N(N-1)(32N^3-268N^2+304N-126)/(N+1)^7 s_2 s_3{}^3\\
&+288N(N-1)(41N^2-50N+23)/(N+1)^6 s_2{}^4 s_3.
\end{aligned}
$$

$$
\begin{aligned}
S(2^2,3,4)={}&N^4(N-1)(N^2-N+1)/(N+1)^{10} s_{11}+44N^3(N-1)(N^2-N+1)/(N+1)^9 s_2 s_9\\
&+2N^2(N-1)(71N^3-130N^2+109N-38)/(N+1)^9 s_3 s_8\\
&+2N(N-1)(157N^4-262N^3+325N^2-126N+24)/(N+1)^9 s_4 s_7\\
&+4N(N-1)(153N^4-289N^3+277N^2-124N+30)/(N+1)^9 s_5 s_6\\
&+4N^2(N-1)(200N^2-149N+140)/(N+1)^8 s_2{}^2 s_7\\
&+48N(N-1)(72N^3-119N^2+89N-21)/(N+1)^8 s_2 s_3 s_6\\
&+8N(N-1)(692N^3-1055N^2+1118N-336)/(N+1)^8 s_2 s_4 s_5\\
&+24N(N-1)(173N^3-329N^2+272N-100)/(N+1)^8 s_3{}^2 s_5\\
&+24N(N-1)(190N^3-429N^2+495N-200)/(N+1)^8 s_3 s_4{}^2\\
&+144N(N-1)(33N^2-20N+14)/(N+1)^7 s_2{}^3 s_5\\
&+48N(N-1)(361N^2-394N+223)/(N+1)^7 s_2{}^2 s_3 s_4\\
&+288N(N-1)(29N^2-68N+40)/(N+1)^7 s_2 s_3{}^3+4320N(N-1)(2N-1)/(N+1)^6 s_2{}^4 s_3.
\end{aligned}
$$

$$
\begin{aligned}
S(2,3^3)={}&N^4(N-1)^3/(N+1)^{10} s_{11}+45N^3(N-1)^3/(N+1)^9 s_2 s_9+9N^2(N-1)^3(17N-9)/(N+1)^9 s_3 s_8\\
&+27N(N-1)^3(11N^2-9N+2)/(N+1)^9 s_4 s_7+9N(N-1)^3(49N^2-36N+12)/(N+1)^9 s_5 s_6\\
&+54N^2(N-1)^2(12N-11)/(N+1)^8 s_2{}^2 s_7+54N(N-1)^2(63N^2-62N+21)/(N+1)^8 s_2 s_3 s_6\\
&+54N(N-1)^2(63N^2-94N+21)/(N+1)^8 s_2 s_4 s_5\\
&+54N(N-1)(46N^3-137N^2+146N-51)/(N+1)^8 s_3{}^2 s_5\\
&+54N(N-1)(82N^3-268N^2+304N-126)/(N+1)^8 s_3 s_4{}^2\\
&+108N(N-1)(30N^2-53N+21)/(N+1)^7 s_2{}^3 s_5\\
&+648N(N-1)(29N^2-68N+40)/(N+1)^7 s_2 s_3 s_4\\
&+648N(N-1)(13N^2-36N+27)/(N+1)^7 s_2 s_3{}^2+1296N(N-1)(5N-7)/(N+1)^6 s_2{}^4 s_3.
\end{aligned}
$$

Weight 12

$$
\begin{aligned}
S(4^3)={}&N^3(N^2-N+1)^3/(N+1)^{11} s_{12}+6N^2(N^2-N+1)^2(11N^2-8N+8)/(N+1)^{10} s_2 s_{10}\\
&+16N(N-1)(N^2-N+1)^2(13N^2-13N+4)/(N+1)^{10} s_3 s_9\\
&+12N(N^2-N+1)(41N^5-115N^4+222N^3-196N^2+113N-24)/(N+1)^{10} s_4 s_8\\
&+48N(N-1)(N^2-N+1)(16N^4-32N^3+54N^2-47N+13)/(N+1)^{10} s_5 s_7\\
&+6N(77N^7-294N^6+717N^5-1088N^4+1233N^3-867N^2+374N-66)/(N+1)^{10} s_6{}^2\\
&+36N(N^2-N+1)(41N^4-60N^3+82N^2-38N+16)/(N+1)^9 s_2{}^2 s_8\\
&+144N(N-1)(N^2-N+1)(51N^3-89N^2+86N-44)/(N+1)^9 s_2 s_3 s_7\\
&+24N(574N^6-1968N^5+4515N^4-5788N^3+5223N^2-2622N+640)/(N+1)^9 s_2 s_4 s_6\\
&+72N(N-1)(109N^5-298N^4+580N^3-692N^2+449N-140)/(N+1)^9 s_2 s_5{}^2\\
&+72N(N-1)(115N^5-444N^4+836N^3-923N^2+630N-202)/(N+1)^9 s_3{}^2 s_6\\
&+288N(N-1)(88N^5-336N^4+716N^3-887N^2+621N-196)/(N+1)^9 s_3 s_4 s_5\\
&+8N(709N^6-3153N^5+8759N^4-13468N^3+13295N^2-7176N+1735)/(N+1)^9 s_4{}^3\\
&+72N(159N^5-378N^4+672N^3-594N^2+396N-96)/(N+1)^8 s_2{}^3 s_6\\
&+864N(N-1)(85N^4-213N^3+304N^2-258N+88)/(N+1)^8 s_2{}^2 s_3 s_5\\
&+72N(709N^5-2132N^4+4614N^3-4802N^2+3082N-762)/(N+1)^8 s_2{}^2 s_4{}^2\\
&+432N(N-1)(276N^4-968N^3+1627N^2-1387N+482)/(N+1)^8 s_2 s_3{}^2 s_4\\
&+72N(N-1)(173N^4-811N^3+1491N^2-1273N+452)/(N+1)^8 s_3{}^4\\
&+864N(57N^4-101N^3+166N^2-94N+29)/(N+1)^7 s_2{}^4 s_4\\
&+864N(N-1)(128N^3-292N^2+269N-106)/(N+1)^7 s_2{}^3 s_3{}^3\\
&+864N(11N^3-8N^2+10N-2)/(N+1)^6 s_2{}^6.
\end{aligned}
$$

$$
\begin{aligned}
S(3^4) = {} & N^4(N-1)^4/(N+1)^{11} s_{12} + 54N^3(N-1)^4/(N+1)^{10} s_2 s_{10} \\
& + 108N^2(N-1)^4(2N-1)/(N+1)^{10} s_3 s_9 + 27N(N-1)^4(17N^4-15N+3)/(N+1)^{10} s_4 s_8 \\
& + 108N(N-1)^4(7N^2-6N+3)/(N+1)^{10} s_5 s_7 + 27N(N-1)^4(17N^4-13N+9)/(N+1)^{10} s_6^2 \\
& + 27N^2(N-1)^3(37N-33)/(N+1)^9 s_2^2 s_8 + 324N(N-1)^3(19N^2-29N+6)/(N+1)^9 s_2 s_3 s_7 \\
& + 162N(N-1)^3(65N^2-118N+45)/(N+1)^9 s_2 s_4 s_6 \\
& + 108N(N-1)^3(59N^2-102N+63)/(N+1)^9 s_2 s_5^2 \\
& + 108N(N-1)^2(82N^3-235N^2+242N-81)/(N+1)^9 s_3^2 s_6 \\
& + 324N(N-1)^2(75N^3-248N^2+295N-138)/(N+1)^9 s_3 s_4 s_5 \\
& + 27N(N-1)(173N^4-811N^3+1491N^2-1273N+452)/(N+1)^9 s_4^3 \\
& + 108N(N-1)^2(71N^2-121N+42)/(N+1)^8 s_2^3 s_6 \\
& + 648N(N-1)^2(79N^2-185N+114)/(N+1)^8 s_2^2 s_3 s_5 \\
& + 486N(N-1)^2(63N^2-164N+125)/(N+1)^8 s_2^2 s_4^2 \\
& + 972N(N-1)(99N^3-391N^2+533N-257)/(N+1)^8 s_2 s_3^2 s_4 \\
& + 162N(N-1)(87N^3-333N^2+493N-263)/(N+1)^8 s_3^4 \\
& + 972N(N-1)(23N^2-57N+38)/(N+1)^7 s_2^4 s_4 \\
& + 648N(N-1)(103N^2-304N+233)/(N+1)^7 s_2^3 s_3^2 + 648N(N-1)(5N-7)/(N+1)^6 s_2^6.
\end{aligned}
$$

$$
\begin{aligned}
S(2^6) = {} & N^6/(N+1)^{11} s_{12} + 60N^5/(N+1)^{10} s_2 s_{10} + 160N^4(N-1)/(N+1)^{10} s_3 s_9 \\
& + 240N^3(2N^2-N+1)/(N+1)^{10} s_4 s_8 + 96N^2(N-1)(7N^2+2)/(N+1)^{10} s_5 s_7 \\
& + 4N(113N^4-68N^3+68N^2-8N+8)/(N+1)^{10} s_6^2 + 1200N^4/(N+1)^9 s_2^2 s_8 \\
& + 4800N^3(N-1)/(N+1)^9 s_2 s_3 s_7 + 2400N^2(7N^2-2N+2)/(N+1)^9 s_2 s_4 s_6 \\
& + 960N(N-1)(6N^2+1)/(N+1)^9 s_2 s_5^2 + 160N^2(N-1)(31N-22)/(N+1)^9 s_3^2 s_6 \\
& + 1920N(N-1)(9N^2-5N+2)/(N+1)^9 s_3 s_4 s_5 + 480N(11N^3-8N^2+10N-2)/(N+1)^9 s_4^3 \\
& + 9600N^3/(N+1)^8 s_2^3 s_6 + 38400N^2(N-1)/(N+1)^8 s_2^2 s_3 s_5 \\
& + 9600N(4N^2-N+1)/(N+1)^8 s_2^2 s_4^2 + 2880N(N-1)(2N-1)/(N+1)^8 s_2 s_3^2 s_4 \\
& + 960N(N-1)(5N-7)/(N+1)^8 s_3^4 + 40320N^2/(N+1)^7 s_2^4 s_4 \\
& + 38400N(N-1)/(N+1)^7 s_2^3 s_3^2 + 3840N/(N+1)^6 s_2^6.
\end{aligned}
$$

Normal Population

$S(4^4) = 6912N(272N^4-379N^3+594N^2-304N+89)/(N+1)^8 s_2^8$ *.

$S(3^6) = 3265920N(N-1)(4N^2-13N+11)/(N+1)^9 s_2^9$ †.

* See footnote on p. 103.

† Stated to have been previously determined.

A PRELIMINARY CLASSIFICATION OF ASIATIC RACES BASED ON CRANIAL MEASUREMENTS.

By T. L. WOO, Ph.D. and G. M. MORANT, D.Sc.

(1) *Introduction.* In a paper published in *Biometrika* in 1921* Miss M. L. Tildesley described a series of Burmese skulls and previously published data relating to other Oriental series were cited. Three years later a paper by one of the present writers in the same *Journal* gave original measurements of a Nepalese and of a Tibetan series and the comparative material used was extended considerably. Comparisons by the method of the coefficient of racial likeness led to results which appeared to be suggestive, but it was clear that nothing approaching even a preliminary classification of Asiatic races could be obtained by these means while the samples were so few in number and, generally, so small in size. Since 1924 the measurements have been published of several new Asiatic cranial series. Professor Harrower's paper of 1926 deals with Southern Chinese and Tamil skulls†, Professor Black's of 1928 with Northern and Prehistoric Chinese, and Dr von Bonin's of 1931 with two Javanese, two Filipino, one Dayak and one Andamanese series. The new data incorporated in these three studies are ample enough to warrant a new survey of the craniology of Asia. All the latest material has been dealt with on biometric lines in the papers cited and use is made below, without further acknowledgment, of a number of statistical constants previously published. The majority of those which are not original are taken from Dr von Bonin's paper and we were greatly indebted to him for permission to use his results before their publication. In his paper a fairly complete comparison has been made on biometric lines of the best Oriental series at present published. Our purpose has been to extend that examination to include all the best series available for the whole of the continent. There is only one good Asiatic cranial series which we have omitted consciously: this is one of Armenians recently described by Professor Bunak‡, for it is clear that the type has closer affinities to some from the south-east of Europe than to any known Asiatic ones. On the other hand, we have included two non-Asiatic series: one is from the Aleutian Islands and the other is composed principally of Kalmucks from Astrakhan. Both these are closely allied to Siberian types. Means for several of the series were obtained by pooling the measurements given by different craniometricians. Care was taken to use these data with as much

* References to all the papers mentioned here will be found in the following section of this paper.

† We have not used the series of Hylam Chinese described by Professor Harrower in *Biometrika*, Vol. xx^{B}, as it is suspected that the crania were artificially distorted.

‡ *Crania Armenica*, Moscow, 1927. This is a supplement to the *Journal Russe d'Anthropologie*, Vol. xvi.

accuracy as possible and many doubtfully defined measurements had to be rejected. Some of the previously published pooled means have been revised and others are now given for the first time. A total of 26 male series was collected and the shortest of these is made up by 31 crania. Nearly all the means used are based on 20 or more crania and those which could only be given for fewer than 10 specimens were ignored. This sample of racial types cannot be supposed a random one taken from all which it may be possible to distinguish among Asiatic peoples. More than half of the series come from Oriental regions; India and Siberia are poorly represented and there are no data whatever referring to the peoples of the south-west of the continent.

The main purpose of the present paper is to present the coefficients of racial likeness between all pairs of the 26 series available. With the usual notation, the form of the crude coefficient used is

$$\frac{1}{M}\Sigma\left(\frac{n_s n_{s'}}{n_s + n_{s'}} \times \frac{(m_s - m_{s'})^2}{\sigma_s^2}\right) - 1 \pm \cdot 67449\sqrt{\frac{2}{M}} = \frac{1}{M}\Sigma(\alpha) - 1 \pm \cdot 67449\sqrt{\frac{2}{M}}.$$

The reduced coefficient is defined to be

$$50 \times \frac{\bar{n}_s + \bar{n}_{s'}}{\bar{n}_s \bar{n}_{s'}}\left\{\frac{1}{M}\Sigma(\alpha) - 1\right\} \pm 50 \times \frac{\bar{n}_s + \bar{n}_{s'}}{\bar{n}_s \bar{n}_{s'}} \times \cdot 67449\sqrt{\frac{2}{M}}.$$

The reduced coefficient is designed to give a measure of racial divergence which does not depend upon the sizes of the samples compared, and the classification suggested is derived from its values. The usual 31 characters, or as many of them as are available in the case of a particular comparison, were used in calculating these constants. Crude coefficients are given for all these measurements and for the indices and angles alone, but only those of the first kind were reduced. The standard deviations of the long Egyptian E series of 26th—30th Dynasty crania described by Professor Karl Pearson and Miss A. G. Davin were used throughout*.

(2) *Measured Series of Asiatic Skulls.* Male adult series are the only ones dealt with in the present paper. Experience has shown that comparisons by the method of the coefficient of racial likeness seldom lead to satisfactory results in the case of very small samples. All the series used have the majority of their means based on 25 or more crania. It is generally an advantage to have all the crania assigned to a single racial type measured by the same worker, but in some cases it is only possible to obtain a large enough sample by pooling the measurements provided from different sources. The need for restricting comparisons in the case of a particular measurement to pairs determined by using identically the same technique is, of course, one of prime importance. Definitions of all those used are given in *Biometrika*, Vol. xx^{B}. 1928, pp. 362—364, where they are denoted by the biometric index letters, and the numbers in Martin's *Lehrbuch* are also given. There are 19 absolute and 12 indicial and angular measurements used in computing the coefficients, but in some cases alternatives can be used as the vertical height

* "On the Biometric Constants of the Human Skull." *Biometrika*, Vol. xvi. 1924, pp. 328—363.

from the basion (H) in place of the basio-bregmatic height (H'), or three comparisons between orbital breadths (O_1, O_1' and Lacrymal O_1) may be possible and one is selected. When such a choice can be made, H' is given preference to H, U to Glabella U, Q' to Bregmatic Q' or Broca's Q', Bregmatic Q' to Broca's Q', O_1 to O_1' or Lacrymal O_1, O_1' to Lacrymal O_1, NH' to NH, R or L, G_1 to G_1' and Alveolar $P\angle$ to Prosthion $P\angle$. One index involving these measurements may be

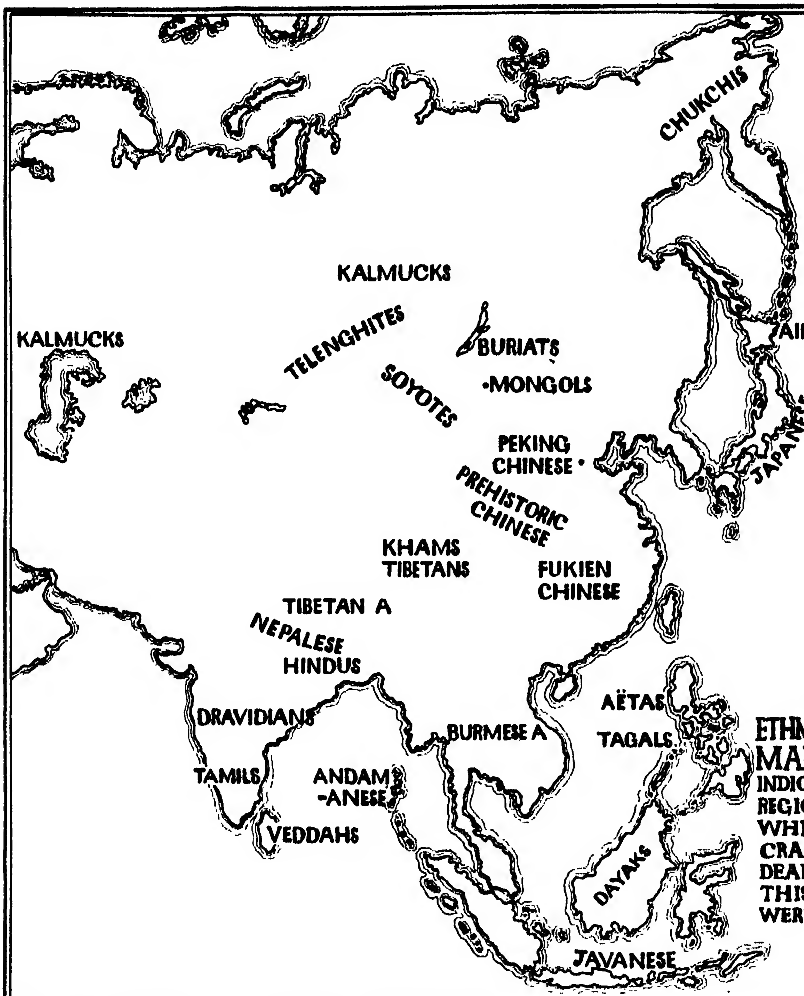


FIG. 1.
ETHNOGRAPHICAL MAP OF ASIA INDICATING THE REGIONS FROM WHICH THE CRANIAL SERIES DEALT WITH IN THIS PAPER WERE OBTAINED

preferred to another as $100\,H'/L$ to $100\,H/L$. The preference in these cases is given to the one which is most frequently available. Most of the alternatives have been provided for the Asiatic series originally described in *Biometrika*. For most of the other series they are not given and several of the coefficients can only be calculated for a smaller number of characters than the total 31. Unless otherwise stated, the means for the series previously published in papers in this *Journal* have been

accepted without modification. Where these were derived from other sources the full references were given and they are not duplicated below. Pooled means which have not previously been published are given in our Table I. The map (p. 110) shows the approximate districts from which the material was derived. The following abbreviations are used in the list below: *A. f. A.* = *Archiv für Anthropologie*; *A. S. D* = *Die anthropologischen Sammlungen Deutschlands*, being supplements to *A. f. A.*; *Bm.* = *Biometrika.*

(i) *Aëtas.* Gerhardt von Bonin: "Beitrag zur Kraniologie von Ost-Asien." *Bm.*, Vol. XXIII. 1931, pp. 52—113. The Aëtas (Philippine Islands) are found in the interior of Luxon Island and also in the islands of Mindoro, Panay and Negros and in the north-east of Mindanao. The tribe is a small one. The people are short of stature and they are generally spoken of as undoubted Negritos. Koeze gave measurements of a series of skulls at Leiden and 33 male specimens were re-measured by von Bonin.

(ii) *Aino.* Koganei's means of a series from Yezo and Kunashiri are quoted in *Bm.*, Vol. I. 1902, p. 426. The calvarial height is the Frankfurt vertical measurement (H) as stated there and not the basio-bregmatic (H') it was assumed to be in some later papers. The index $100\,B/H$ is {101·2 (88)} in place of 98·8. We have omitted the orbital breadth and index and the palatal measurements, as they are inadequately defined, and added means for *fml* (35·7 (76)), *fmb* (30·2 (81)), 100 *fmb*/*fml* (84·6 (76)), $N\angle$ ({70°·2 (69)}) and $A\angle$ ({71°·2 (69)}). The profile angle is assumed to have been measured from the prosthion and not from the alveolar point. There are 88 male skulls.

(iii) *Andamanese.* Pooled means of various short series, of which one has been re-measured by von Bonin, are given by him in *Bm.*, Vol. XXIII. 1931, pp. 84—85. There are 34 male skulls.

(iv) *Aleuts.* Measurements of male Aleutian crania were taken from the following sources:

(*a*) Aleš Hrdlička: "Catalogue of Human Crania in the United States National Museum Collections." *Proceedings of the United States National Museum*, Vol. LXIII. 1924, pp. 1—51 (25 crania).

(*b*) A. Tarenetzky: "Beiträge zur Skelet- und Schädelkunde der Aleuten, Konaegen, Kenai und Koljuschen mit vergleichend anthropologischen Bemerkungen." *Mémoires de l'Académie Impériale des Sciences de St Pétersbourg*, VIII^e série, T. IX. 1900, pp. 1—73 (7 crania).

(*c*) Herman F. C. ten Kate: *Zur Craniologie der Mongoloiden: Beobachtungen und Messungen.* Diss., Berlin, 1882 (1 cranium).

(*d*) George Montandon: "Craniologie Paléosibérienne. Seconde Partie." *L'Anthropologie*, T. XXXVI. 1926, pp. 447—542 (2 crania, omitting No. 4829 which is deformed).

These skulls came from several different islands of the Aleutian archipelago and one from Kodiak Island is included. Pooled means are given in Table I below.

(v) *Buriats.* Measurements of male Buriat crania were taken from the following sources:

(*a*) Aleš Hrdlička: *loc. cit.*, pp. 46—47 (19 crania).

(*b*) Michael Reicher: "Untersuchungen über die Schädelform der alpenländischen und mongolischen Brachycephalen." *Zeitschrift für Morphologie und Anthropologie*, Bd. xv. 1913, Tabelle 3 *a* (15 crania).

(*c*) Julius Fridolin: "Burjäten- und Kalmückenschädel." *A. f. A.*, Bd. xxvii. 1900—1902, S. 304—305 (7 crania).

(*d*) Herman F. C. ten Kate: *op. cit.* (4 crania). A few additional measurements of these specimens which we were able to use are given by Haberkorn in *Zeitschrift für Ethnologie*, Bd. x. 1878, S. 307.

The localities from which the majority of these Buriat skulls were obtained are known and they all lie in a comparatively small area round the southern extremity of Lake Baikal, and nearly all are to the east of that lake. The Buriats are known to have moved into this region from the Amur District north of Manchuria in the 13th century.

(vi) *Burmese A.* A series from the neighbourhood of Moulmein was divided into three groups of which the Burmese *A*, supposed true Burman, is one. Measurements were taken by Miss Tildesley and means are given in *Bm.*, Vol. xiii. 1921, p. 239. In computing the coefficients we have added to these $N\angle$ (66°·8 (38)) and $A\angle$ (70°·5 (38)). The palatal breadth was taken between the inner alveolar walls at the second molars, and this is said to have been less in some cases than the breadth between the inner rims of the alveoli of the second molars. The second of these measurements, which is Martin's and the one generally used to-day, gives a male mean of 41·6 (29), giving indices $100\ G_2/G_1 = 84{\cdot}0$ (27) and $100\ G_2/G_1' =$ {91·6 (29)}. These values were used in computing the coefficients. There are 44 male Burmese *A* skulls.

(vii) *Chinese: Fukien.* Gordon Harrower: "A Study of the Hokien and the Tamil Skull." *Transactions of the Royal Society of Edinburgh*, Vol. lix. Part III (No. 13), 1926, pp. 573—599. Measurements of 36 male skulls of unclaimed coolies from the southern Chinese province of Fukien (Hokien) are given. The means are quoted, with a few corrections, in *Bm.*, Vol. xxiii. 1931, pp. 84—85.

(viii) *Chinese* (*Koganei*). In 1902 Koganei gave measurements of the skulls of 70 Chinese soldiers who had been killed in the war with Japan. The collection was made in the northern provinces of Shantung and Chihli and in Southern Manchuria, but the regions from which the soldiers came are unknown. Means are quoted in *Bm.*, Vol. xvi. 1924, pp. 48—49. In calculating the coefficients we omitted the insufficiently defined palatal measurements and the profile angle was assumed to be from the prosthion and not from the alveolar point.

(ix) *Chinese: Peking.* Davidson Black: "A Study of Kansu and Honan Aeneolithic Skulls and Specimens from later Kansu Prehistoric Sites in Comparison with North China and other recent Crania. Part I. On Measurements and

Identification." *Palaeontologia Sinica*, Series D, Vol. VII. 1928, pp. 1—83. Measurements are given of 86 male skulls from the northern provinces of China collected in Peking dissecting rooms. The majority of the men came from Chihli, Shansi and Shantung, but there were some from Shensi, Fengtien and Northern Honan. Means are quoted in *Bm.*, Vol. XXIII. 1931, pp. 84—85.

(x) *Chinese: Prehistoric.* Davidson Black: *loc. cit.* The Pooled Prehistoric series described in this paper comprises skulls from Kansu and Honan of the Early Bronze, Copper and Aeneolithic periods. It is shown that there is sufficient justification for combining this material. There are 64 male specimens though many of these are imperfect. Means are quoted in *Bm.*, Vol. XXIII. 1931, pp. 84—85.

(xi) *Chukchis.* Julius Fridolin: "Tschuktschenschädel." *A. f. A.*, Bd. XXVIII. Supplement, 1904, S. 1—17. Measurements of 35 male skulls are given. The Chukchis inhabit the extreme north-east of Asia with the exception of some points on the coast which are said to be occupied by Eskimos. Of these specimens described by Fridolin 20 came from the Chukchi area proper and 15 from the Eskimo area. Montandon (*loc. cit. infra*, pp. 284—285) has given the means of 100 B/L, 100 H/L and 100 NB/NH for the two groups separately and also for the corresponding female and juvenile groups. The cephalic indices are practically identical in the case of the adult groups compared, while the Eskimo area group has the higher height-length and the lower nasal indices for both sexes. The difference between the male height-length indices is just significant and all others are insignificant. It is probable that significant differences between the types derived from the two areas would be found if more adequate material were available. In order to obtain a large enough sample for present purposes, however, we pooled all the specimens measured by Fridolin. Male means are given in Table I below and these were used in computing the coefficients with the other Asiatic series. Measurements of other male Chukchi skulls are given in the following sources:

(*a*) Aleš Hrdlička: *loc. cit.*, pp. 16—17 (5 crania, 4 of which may be of mixed Chukchi and Eskimo origin).

(*b*) George Montandon: "Craniologie Paléosibérienne (Néolithiques, Mongoloïdes, Tchouktchi, Eskimo, Aléoutes, Kamtchadales, Aïnou, Ghiliak, Négroïdes du Nord)." *L'Anthropologie*, T. XXXVI. 1926, pp. 209—296 (10 crania, omitting No. 4 which is distorted, of which the majority are Chukchi proper and a few may be of mixed origin).

The pooled cephalic index for Hrdlička's and Montandon's short series is 75·7 (14) and for 9 characters a coefficient is found with Fridolin's series of 2·26 ± ·32 (reduced 11·74 ± 1·65). The pooling of all the material is hardly justified.

(xii) *Dayaks.* Gerhardt von Bonin: *loc. cit.* Original measurements of three short series of Dayak skulls at Leiden were pooled with those of another described by Emil Schmidt. There are 55 of these male skulls from Borneo in all.

(xiii) *Dravidians.* Means of two Indian series measured by Sir William Turner are given by Miss B. N. Stoessiger in *Bm.*, Vol. XIX. 1927, p. 128. The first is of Dravidian skulls from the Central Provinces and Orissa, including one Tamil from Madras, and the second of Dravidians ("Kolarians") from Southern India. An insignificant coefficient was found between these groups and hence they were pooled, giving a total of 32 male specimens. These means are distinctly differentiated from those of a Maravar (Dravidian) series from Madras. The Dravidian series used below is the pooled one derived from Turner's measurements. It was necessary to modify the published means: the horizontal circumference is Glabella U and not U, the orbital breadth Lacrymal O_1 and not O_1', the transverse arc is Broca's Q' and not Q', and the mean should be 296·1 (31) in place of 302·0 (34); the angles of the fundamental triangle are based on 28 skulls, not 30. The capacity was omitted in calculating the coefficients.

(xiv) *Hindus.* Measurements of male Hindu crania were taken from the following sources:

(*a*) Jacopo Danielli: "Studio sui crani bengalesi." *Archivio per l'Antropologia e l'Etnologia*, Vol. XXII. 1892, pp. 371—448. Measurements are given of 42 male crania of Hindus of the inferior castes from the banks of the lower Ganges.

(*b*) Sir William Turner: "Contributions to the Craniology of the People of the Empire of India. Part II. The Aborigines of Chúta Nágpúr and of the Central Provinces, the People of Orissa, the Veddahs and Negritos." *Transactions of the Royal Society of Edinburgh*, Vol. XL. 1901, pp. 59—129. Measurements are given (Tables VI—VIII) of 25 male skulls classed as Uriyá (or Ooriá). This is the mother-tongue of the vast majority of the Hindu peoples of Orissa who inhabit the plains.

(*c*) Paolo Mantegazza: "Studii sull' etnologia dell' India." *Archivio per l'Antropologia e l'Etnologia*, Vol. XIII. 1883, pp. 177—241. Measurements are given (pp. 212—215) of 24 male skulls from Southern India.

Coefficients of racial likeness between these three groups of Hindu skulls are given in *Bm.*, Vol XX^B. 1928, p. 298; two values are insignificant and the other is $1{\cdot}74 \pm {\cdot}21$. Pooled means are given in Table I below.

(xv) *Japanese.* Pooled means, based principally on the measurements of Ono and Adachi, are given in *Bm.*, Vol. XXIII. 1931, pp. 84—85. There are 138 male skulls from different parts of both islands represented, but few means are available for more than 50 specimens.

(xvi) *Javanese: Bantam and Batavia.* Gerhardt von Bonin: *loc. cit.* This series from the west of the island comprises 55 male skulls preserved at Leiden.

(xvii) *Javanese: Middle and East. Ibid.* The 65 male skulls at Leiden on which the means are based came principally from the middle and east of the island, though a few are from unknown localities.

(xviii) *Kalmucks.* Measurements of Kalmuck skulls were taken from the following sources:

(*a*) S. Sommier: "Note di Viaggio. II. Mordvá-Popolazione di Astrakan-Kalmucchi." *Archivio per l'Antropologia e l'Etnologia*, Vol. XIX. 1889, pp. 117—157 (7 crania from Astrakhan).

(*b*) Julius Fridolin: *loc. cit.* (v)*c* (9 crania from Astrakhan Province, 3 from Tomsk Province and 3 from unknown localities). A few additional measurements of these skulls taken by C. Mérejkowsky (*Revue d'Anthropologie*, 2e série, T. VII. 1884, pp. 296—297) could also be used.

(*c*) Michael Reicher: *loc. cit.* (19 crania from Astrakhan Province).

(*d*) Herman F. C. ten Kate: *loc. cit.* (1 cranium from Astrakhan Province and 3 from unknown localities). Additional measurements of 3 of these specimens given by Haberkorn (*loc. cit.*) could also be used.

(*e*) J. W. Spengel: *A. S. D.*, Göttingen Catalogue, 1874, S. 40 (3 crania from the Astrakhan district and 3 from unknown localities).

(*f*) J. Deniker: "Étude sur les Kalmouks. Suite." *Revue d'Anthropologie*, 2e série, T. VII. 1884, pp. 277—310 (4 crania from unknown localities previously measured by Quatrefages and Hamy).

The majority of the skulls forming these short series are known to have been obtained in the Astrakhan region, 3 came from Central Siberia and a certain number from unknown localities. Pooled means are given in Table I below.

(xix) *Mongols.* Aleš Hrdlička: *loc. cit.*, pp. 40—43. There are 114 male skulls from Urga in Northern Mongolia which is immediately to the south of and 250 miles distant from Lake Baikal. The capacities given were not used in calculating the coefficients. Means are quoted in Table I below.

(xx) *Nepalese.* G. M. Morant: "A Study of certain Oriental Series of Crania including the Nepalese and Tibetan Series in the British Museum (Natural History)." *Bm.*, Vol. XVI. 1924, pp. 1—104. There are 48 male skulls from different parts of the country.

(xxi) *Soyotes.* G. Debetz: "The Anthropological Type of the Turanians of the Kemtchik and Tannu Regions (Soyotes)." *North Asia*, 1929, pp. 127—140 (in Russian). The Soyotes (Soyots, Soyons or Soiotes) are presumed to be of Finno-Turkic origin. They form a small community to-day in the border country between the Sayan and Altai mountains to the west of Lake Baikal. Means given for 9 male skulls from the Kemtchik and 31 from the central region are closely similar and the pooled values are given in Table I below.

(xxii) *Tagals.* Gerhardt von Bonin: *loc. cit.* The Tagals (Tagálogs) of the Philippine Islands constitute the bulk of the population of Manila, Mindanao and central Luxon. They are classed as one of the brown, non-negrito tribes of the islands. Koeze gave measurements of a cranial series at Leiden and 31 male specimens were re-measured by von Bonin.

(xxiii) *Tamils.* Gordon Harrower: *loc. cit.* Unclaimed bodies of 35 Tamil coolies were available for study at Singapore. Apart from the fact that the men came from Southern India, nothing is known of their origin. The means are quoted —a few being corrected—in Table I below.

(xxiv) *Telenghites.* The Telenghites (or Teleuts) are a small Tatar tribe inhabiting to-day the lowlands of the Altai region of Southern Siberia between Lakes Balkash and Baikal. Means reduced from Reicher's measurements of 60 male skulls are given in *Bm.*, Vol. XXIII. 1931, pp. 84—85.

(xxv) *Tibetans A.* G. M. Morant: *loc. cit.* Means are given for 37 male skulls from the south-west of the country which conform, as all the inhabitants of that district are supposed to do, to the Tibetan *A* type.

(xxvi) *Veddahs.* Measurements of male Veddah skulls were taken from the following sources:

(*a*) Sir William Turner: *loc. cit.* (7 crania).

(*b*) Paul u. Fritz Sarasin: *Ergebnisse naturwissenschaftlicher Forschungen auf Ceylon*, Bd. III. Wiesbaden, 1892—3, S. 198—307 (22 crania). Most of these skulls were subsequently measured by Lüthy (*A.f. A.*, Bd. XXXIX. 1912, S. 70) and a few of his additional measurements can be used.

(*c*) Arthur Thomson: "On the Osteology of the Veddahs of Ceylon." *Journal of the Anthropological Institute*, Vol. XIX. 1889, pp. 125—158 (6 crania at Oxford).

(*d*) William Henry Flower: *Catalogue of the Specimens...in the Museum of the Royal College of Surgeons, Part I.* 1907 (15 crania).

(*e*) Rudolf Virchow: "Ueber die Veddahs von Ceylon." *Abhandlungen der Königl. Akademie der Wissenschaften zu Berlin*, 1881 (1 cranium).

(*f*) Rudolf Virchow: *Zeitschrift für Ethnologie*, Bd. XIV. 1882, Verhandlungen, S. 302, and Bd. XVII. 1885, Verhandlungen, S. 500 (3 crania).

(*g*) Rüdinger: *A. S. D.*, München Catalogue, 1892, No. 414c (1 cranium).

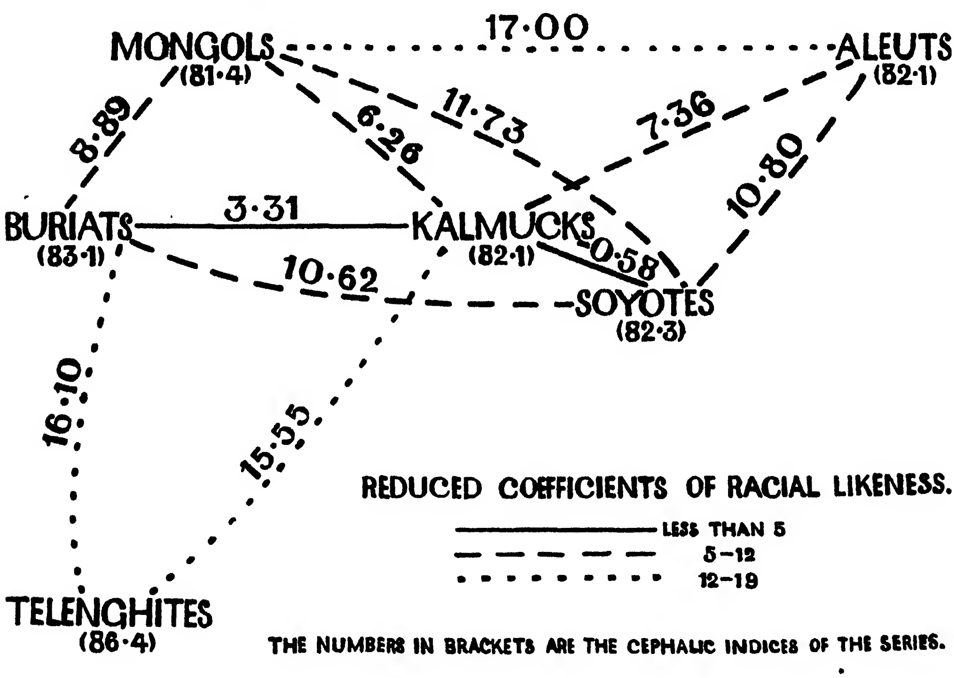
The pooled means for these Veddah skulls are given in Table I below.

These 26 series have been used in order to arrive at a preliminary classification of the races they represent, and for this purpose the coefficients of racial likeness were calculated for every possible pair. A few coefficients have also been found with the following shorter series which represents a particularly interesting racial type.

(xxvii) G. M. Morant: "A First Study of the Tibetan Skull." *Bm.*, Vol. XIV. 1923, pp. 193—260. The series comprises 15 male skulls from the eastern province of Khams said to belong to the Tibetan *B* type which is clearly differentiated from the Tibetan *A*.

(3) *Coefficients of Racial Likeness between Asiatic Series.* When the coefficients of racial likeness between all the Asiatic series are compared, it is found that a sharp distinction is made between a group of six which are closely allied to one another and all the remaining ones (cf. Tables II, V and VI below). The coefficients between these six are given in Table II. Some of the regions represented are widely separated geographically. The vast majority of the Kalmuck skulls used come from the Astrakhan region which is 2000 miles to the west of the district inhabited by the neighbouring Telenghites (Teleuts), Buriats, Soyotes and Mongols; and it is more than 3000 miles from there to the Aleutian Islands. But all the races in the group

FIG. 2. THE RELATIONSHIPS OF NORTHERN MONGOLIAN RACES.



come from the northern parts of Asia and the islands in the Bering Sea, while the Chukchi is the only other type from that region for which we have adequate craniological data. The group may be referred to as that of the Northern Mongolian races, and it is probable that it would also include the Samoyed, Tungus, Ostiaks and Yakuts. The scheme of relationship suggested by the lowest reduced coefficients is shown in Fig. 2. It is surprising to find that the samples representing the Kalmucks and Soyotes may be supposed to represent identically the same racial type. In spite of the fact that no distinction can be made between these two when they are compared directly, their relationships to the other series are by no means the same. The Buriats are very close to the Kalmucks, but distinctly removed from

TABLE I.

Mean Measurements of Male Series of Asiatic Crania.*

Character	Aleuts (pooled)	Buriats (pooled)	Kalmucks (pooled)	Soyotes (Debetz)	Mongols (Hrdlička)	Chukchis (Fridolin)	Hindus (pooled)	Tamils (Harrower)	Veddahs (pooled)
C	—	—	—	—	1573·0 (112)	—	—	1350·3 (35)	—
F	—	—	—	—	—	—	174·9 (42)	—	178·8 (21)
L	181·0 (35)	179·9 (45)	181·5 (55)	183·3 (40)	184·0 (113)	181·7 (35)	176·3 (91)	179·6 (35)	178·6 (39)
B	148·7 (35)	150·0 (45)	149·3 (55)	150·8 (40)	150·1 (113)	141·2 (35)	132·3 (91)	131·5 (35)	128·5 (54)
B'	92·4 (9)	96·3 (19)	96·3 (40)	95·8 (39)	—	—	92·1 (91)	95·2 (35)	91·0 (36)
B''	120·5 (2)	124·8 (14)	121·7 (19)	—	—	—	111·6 (39)	—	107·9 (19)
Biasterionic B	—	112·9 (15)	115·2 (19)	—	—	—	102·6 (65)	103·7 (35)	100·6 (16)
H'	127·6 (27)	131·1 (37)	129·3 (47)	130·7 (39)	131·0 (111)	—	133·3 (91)	136·3 (35)	132·6 (43)
H	—	133·0 (7)	130·2 (21)	—	—	135·1 (34)	—	—	133·2 (26)
LB	97·6 (10)	101·6 (26)	100·6 (44)	101·4 (39)	—	101·8 (34)	98·3 (67)	101·5 (35)	97·8 (48)
S_1'	—	113·8 (15)	113·1 (19)	—	—	—	—	112·2 (35)	—
S_2'	—	107·9 (14)	108·9 (19)	—	—	—	—	—	—
S_3'	—	92·8 (18)	94·7 (22)	—	—	—	—	92·4 (35)	—
S_1	123·5 (6)	127·2 (19)	126·8 (34)	—	—	—	126·7 (91)	129·4 (35)	129·5 (16)
S_2	115·7 (6)	122·1 (17)	119·2 (34)	—	—	—	124·0 (87)	129·8 (35)	129·3 (16)
S_3	116·1 (8)	111·6 (18)	115·3 (34)	—	—	—	110·3 (86)	109·0 (35)	109·1 (33)
S	354·5 (10)	360·6 (24)	361·7 (51)	—	—	362·8 (35)	361·2 (89)	367·6 (35)	364·7 (35)
U	521·3 (3)	524·9 (26)	524·4 (55)	—	—	511·9 (34)	489·0 (66)	498·0 (35)	492·0 (16)
Glabella U	516·2 (5)	530·4 (15)	525·7 (19)	—	—	—	495·7 (25)	—	498·1 (16)
Q'	—	316·1 (7)	314·1 (15)	—	—	312·8 (35)	—	313·2 (35)	299·5 (23)
Bregmatic Q'	—	324·2 (15)	310·9 (19)	—	—	—	—	—	—
Broca's Q'	302·0 (3)	314·5 (4)	307·1 (11)	—	—	—	295·9 (91)	—	292·7 (13)
fml	35·9 (9)	38·2 (19)	37·8 (22)	—	—	—	33·9 (66)	36·7 (35)	34·6 (15)
fmb	30·5 (9)	31·3 (19)	31·5 (22)	—	—	—	29·3 (42)	29·5 (35)	30·0 (2)
$G'H$	74·1 (26)	76·1 (33)	74·8 (40)	75·2 (29)	77·6 (79)	76·5 (31)	66·4 (61)	64·8 (35)	61·3 (28)
GB	—	102·6 (18)	103·0 (31)	—	—	103·0 (32)	—	97·5 (35)	91·8 (21)
J	144·0 (33)	139·8 (37)	140·5 (51)	140·7 (31)	142·0 (102)	139·5 (31)	124·3 (79)	127·8 (35)	124·4 (34)
GL	102·0 (3)	96·7 (8)	100·9 (16)	—	—	—	96·8 (23)	96·4 (35)	93·2 (39)
NH'	52·7 (35)	55·4 (37)	54·4 (45)	54·2 (34)	56·6 (114)	—	48·5 (88)	49·4 (35)	45·0 (23)
NH, R or L	—	53·6 (7)	56·3 (14)	—	—	53·5 (35)	—	—	47·3 (22)

TABLE I (*continued*).

Character	Aleuts (pooled)	Buriats (pooled)	Kalmucks (pooled)	Soyotes (Debetz)	Mongols (Hrdlička)	Chukchis (Fridolin)	Hindus (pooled)	Tamils (Harrower)	Veddahs (pooled)
NB	25·5 (35)	27·1 (42)	26·7 (45)	26·8 (33)	27·5 (114)	25·4 (35)	24·5 (88)	24·6 (35)	24·2 (47)
O_1, R or L	—	42·6 (15)	41·9 (19)	—	—	—	—	—	—
O_1', R or L	40·3 (25)	37·6 (18)	—	40·6 (36)	39·3 (112)	—	—	39·5 (35)	—
Lacrymal O_1, R or L	40·5 (10)	39·3 (4)	40·7 (27)	—	—	—	37·6 (85)	—	37·6 (43)
O_2, R or L	36·2 (34)	34·9 (44)	35·3 (48)	34·7 (36)	35·9 (112)	35·9 (35)	32·6 (85)	32·5 (35)	33·3 (46)
SC	—	7·6 (15)	8·5 (18)	—	—	—	8·6 (39)	—	7·8 (19)
DC	—	—	—	23·0 (34)	—	—	—	—	—
G_1	—	—	—	—	—	—	—	50·2 (35)	—
G_1'	—	46·7 (9)	46·4 (18)	—	—	—	—	—	—
G_2	—	43·3 (8)	42·9 (15)	—	—	—	—	38·8 (35)	—
$100\ B/L$	82·1 (35)	83·1 (45)	82·1 (55)	82·3 (40)	81·4 (113)	77·7 (35)	75·1 (91)	73·3 (35)	72·0 (39)
$100\ H'/L$	{70·0 (27)}	{72·9 (37)}	71·0 (47)	71·3 (39)	{71·2 (111)}	—	75·7 (91)	76·2 (35)	73·4 (28)
$100\ H/L$	—	74·7 (7)	71·4 (21)	—	—	74·3 (34)	—	—	74·7 (26)
$100\ B/H'$	{118·0 (27)}	{113·6 (37)}	{115·5 (47)}	{115·4 (39)}	{114·6 (111)}	—	{99·2 (91)}	96·1 (35)	{96·9 (43)}
$100\ B/H$	—	111·6 (7)	{114·7 (21)}	—	—	{104·5 (34)}	—	—	{96·5 (26)}
$100\ fmb/fml$	85·0 (9)	82·4 (15)	83·4 (22)	—	—	—	{85·0 (42)}	80·2 (35)	80·1 (2)
$Oc.\ I.$	—	60·2 (18)	59·2 (22)	—	—	—	—	62·2 (35)	—
$100\ G'H/GB$	—	74·5 (15)	72·6 (29)	—	—	74·3 (30)	—	66·7 (35)	69·3 (15)
$100\ NB/NH'$	48·4 (35)	48·8 (37)	{49·2 (45)}	49·6 (33)	48·6 (114)	—	50·6 (88)	51·6 (35)	52·9 (23)
$100\ NB/NH$	—	51·8 (7)	47·0 (14)	—	—	47·6 (35)	—	—	52·2 (22)
$100\ O_2/O_1$	—	84·3 (15)	81·9 (19)	—	—	—	—	—	—
$100\ O_2/O_1'$	89·5 (24)	91·0 (18)	—	85·2 (36)	91·5 (112)	—	—	82·2 (35)	—
100 O_2/Lacrymal O_1	90·0 (10)	90·7 (4)	{86·7 (27)}	—	—	—	86·8 (85)	—	87·7 (43)
$100\ G_2/G_1$	—	—	—	—	—	—	—	79·7 (35)	—
$100\ G_2/G_1'$	—	92·1 (8)	91·5 (15)	—	—	—	—	—	—
Alveolar $P\angle$	—	—	—	—	—	—	—	85°·5 (35)	—
Prosthion $P\angle$	—	87°·4 (11)	85°·4 (15)	86°·7 (31)	—	—	—	—	88°·1 (1)
$N\angle$	71°·9 (3)	64°·9 (8)	{66°·9 (16)}	—	—	—	{68°·0 (23)}	66°·4 (35)	{67°·0 (28)}
$A\angle$	66°·3 (3)	70°·7 (8)	{67°·8 (16)}	—	—	—	{73°·5 (23)}	75°·5 (35)	{75°·8 (28)}
$B\angle$	41°·8 (3)	44°·3 (8)	{45°·3 (16)}	—	—	—	{38°·5 (23)}	38°·1 (35)	{37°·2 (28)}

* The means in this table have not been given before in *Biometrika* and those of the other series used in the present paper will be found in earlier volumes of this *Journal*. The capacities were omitted in the case of the pooled series as this measurement has been determined in a number of ways which may give sensibly different results. The mean indices and angles in curled brackets were found from the means of component absolute measurements instead of from individual values.

TABLE II.

*Coefficients of Racial Likeness between Northern Mongolian Races**.

		Crude Coefficients					
		Telenghites (55·4)	Buriats (28·5)	Kalmucks (35·8)	Soyotes (36·1)	Mongols (109·5)	Aleuts (28·3)
Telenghites (55·4)	All characters ...	—	6·21 ± ·19 (24)	6·76 ± ·18 (28)	9·61 ± ·25 (15)	38·79 ± ·28 (12)	12·40 ± ·25 (14)
	Indices and Angles		6·73 ± ·32 (9)	9·61 ± ·30 (10)	15·80 ± ·43 (5)	47·80 ± ·48 (4)	24·47 ± ·48 (4)
Buriats (28·5)	All characters ...	6·21 ± ·19 (24)	—	1·08 ± ·19 (25)	3·69 ± ·23 (17)	4·88 ± ·25 (14)	7·84 ± ·24 (16)
	Indices and Angles	6·73 ± ·32 (9)		1·46 ± ·32 (9)	3·57 ± ·39 (6)	4·78 ± ·43 (5)	6·48 ± ·43 (5)
Kalmucks (35·8)	All characters ...	6·76 ± ·18 (28)	1·08 ± ·19 (25)	—	−0·23 ± ·25 (15)	4·19 ± ·28 (12)	2·47 ± ·24 (16)
	Indices and Angles	9·61 ± ·30 (10)	1·46 ± ·32 (9)		−0·65 ± ·43 (5)	0·61 ± ·48 (4)	1·56 ± ·43 (5)
Soyotes (36·1)	All characters ...	9·61 ± ·25 (15)	3·69 ± ·23 (17)	−0·23 ± ·25 (15)	—	6·36 ± ·25 (14)	3·52 ± ·25 (15)
	Indices and Angles	15·80 ± ·43 (5)	3·57 ± ·39 (6)	−0·65 ± ·43 (5)		7·79 ± ·43 (5)	2·86 ± ·43 (5)
Mongols (109·5)	All characters ...	38·79 ± ·28 (12)	4·88 ± ·25 (14)	4·19 ± ·28 (12)	6·36 ± ·25 (14)	—	8·20 ± ·25 (14)
	Indices and Angles	47·80 ± ·48 (4)	4·78 ± ·43 (5)	0·61 ± ·48 (4)	7·79 ± ·43 (5)		3·45 ± ·43 (5)
Aleuts (28·3)	All characters ...	12·40 ± ·25 (14)	7·84 ± ·24 (16)	2·47 ± ·24 (16)	3·52 ± ·25 (15)	8·20 ± ·25 (14)	—
	Indices and Angles	24·47 ± ·48 (4)	6·48 ± ·43 (5)	1·56 ± ·43 (5)	2·86 ± ·43 (5)	3·45 ± ·43 (5)	
		Reduced Coefficients (all characters)					
Telenghites ...	... ...	—	16·10 ± ·50	15·55 ± ·41	21·80 ± ·56	51·07 ± ·36	32·03 ± ·66
Buriats ...	... ...	16·10 ± ·50	—	3·31 ± ·58	10·62 ± ·67	8·89 ± ·45	25·01 ± ·76
Kalmucks ...	... ...	15·55 ± ·41	3·31 ± ·58	—	−0·58 ± ·61	6·26 ± ·41	7·36 ± ·71
Soyotes ...	... ...	21·80 ± ·56	10·62 ± ·67	−0·58 ± ·61	—	11·73 ± ·47	10·80 ± ·76
Mongols ...	... ...	51·07 ± ·36	8·89 ± ·45	6·26 ± ·41	11·73 ± ·47	—	17·00 ± ·53
Aleuts ...	... ...	32·03 ± ·66	25·01 ± ·76	7·36 ± ·71	10·80 ± ·76	17·00 ± ·53	—

* The numbers following the designation of the race are the mean numbers of skulls available for the characters used in computing the coefficients—the $\bar{n}$'s—in the case of the comparison which involves the largest number of these characters for the particular series. When fewer than this maximum number can be used the $\bar{n}$ may differ to some extent from the one given. The numbers in brackets following the coefficients are the numbers of characters on which they are based.

TABLE III.

*Coefficients of Racial Likeness between Indian Races**.

		Crude Coefficients				
		Veddahs (34·2)	Dravidians (31·1)	Tamils (35·0)	Hindus (76·1)	Nepalese (45·6)
Veddahs (34·2)	All characters ...	—	3·36 ± ·20 (22)	7·35 ± ·21 (21)	6·53 ± ·20 (22)	6·70 ± ·21 (21)
	Indices and Angles		3·79 ± ·36 (7)	2·62 ± ·36 (7)	9·34 ± ·36 (7)	7·27 ± ·36 (7)
Dravidians (31·1)	All characters ...	3·36 ± ·20 (22)	—	4·98 ± ·22 (18)	6·51 ± ·20 (22)	6·87 ± ·24 (16)
	Indices and Angles	3·79 ± ·36 (7)		5·04 ± ·39 (6)	10·38 ± ·36 (7)	8·23 ± ·43 (5)
Tamils (35·0)	All characters ...	7·35 ± ·21 (21)	4·98 ± ·22 (18)	—	6·63 ± ·21 (21)	6·42 ± ·18 (29)
	Indices and Angles	2·62 ± ·36 (7)	5·04 ± ·39 (6)		5·63 ± ·36 (7)	5·56 ± ·29 (11)
Hindus (76·1)	All characters ...	6·53 ± ·20 (22)	6·51 ± ·20 (22)	6·63 ± ·21 (21)	—	3·05 ± ·22 (19)
	Indices and Angles	9·34 ± ·36 (7)	10·38 ± ·36 (7)	5·63 ± ·36 (7)		0·51 ± ·39 (6)
Nepalese (45·6)	All characters ...	6·70 ± ·21 (21)	6·87 ± ·24 (16)	6·42 ± ·18 (29)	3·05 ± ·22 (19)	—
	Indices and Angles	7·27 ± ·36 (7)	8·23 ± ·43 (5)	5·56 ± ·29 (11)	0·51 ± ·39 (6)	
		Reduced Coefficients (all characters)				
Veddahs ...	... ...	—	10·32 ± ·62	21·66 ± ·61	13·85 ± ·43	17·51 ± ·54
Dravidians	... ...	10·32 ± ·62	—	15·15 ± ·68	14·77 ± ·46	18·58 ± ·65
Tamils ...	... ...	21·66 ± ·61	15·15 ± ·68	—	13·97 ± ·44	16·21 ± ·45
Hindus ...	... ...	13·85 ± ·43	14·77 ± ·46	13·97 ± ·44	—	5·43 ± ·39
Nepalese ...	... ...	17·51 ± ·54	18·58 ± ·65	16·21 ± ·45	5·43 ± ·39	—

* See footnote to Table II.

the Soyotes. Each one of the five other series has its lowest reduced coefficient with the Kalmucks, but the very similar Soyotes do not occupy such a central position. The Telenghites are distinguished from the other series by having no close connection with any one of them. There seems to be little relation between the affinities of the types and their geographical positions, and this is doubtless due to the fact that most of them represent nomadic, or semi-nomadic, peoples. The same condition would doubtless explain why several pairs of widely separated races have such close relationships, but far more material might be needed to establish any close correspondence between the known migrations and the present-day affinities of the types.

The second group of Asiatic races distinguished by the coefficients is one made up by all the Indian series available. Crude and reduced values are given in Table III and the arrangement suggested by the latter is shown in Fig. 3. The connections are less close than those between the Northern Mongolian races, there being no reduced coefficient less than 5 and only two—Hindus with Nepalese and Veddahs with Dravidians—less than 12. But when a rather higher limit is considered, the series are found to have numerous connections with one another. Every one has a reduced coefficient less than 19 with every other one, except in the case of the Veddahs and Tamils, and there the value is 21·66. It is interesting to note that the Veddahs and Dravidians have their lowest coefficient with one another, while no sharp distinction can be made between them and the other Indian races. There is a rough correspondence between the affinities and geographical positions of the races, though an exception to this is the fact that the Tamils from Southern India approach most closely to the Hindus of Bengal, while they stand appreciably closer to the Nepalese than to the Veddahs of Ceylon.

The coefficients of racial likeness between 12 of the remaining series are given in Table IV and the connections provided by the reduced values less than 19 are shown in Fig. 3. A number of low values are found and every one of these types has at least one reduced coefficient less than 10. The group as a whole will be referred to as that of the Oriental races, and it can be distinguished clearly from both the Northern Mongolian and Indian groups (see Tables VI and VII). In spite of the large distances which separate the localities from which the three modern Chinese series and the Japanese were procured, the types are found to be closely similar*. The Prehistoric series is equally and more distantly related to the three others from China, but this connection is still more intimate than any which has been found between one or another of the Veddah, Dravidian, Tamil or Telenghite groups and any other Asiatic series. The races of China represented here and the Japanese may be considered to form a sub-group of the Oriental one, their closest relationships to one another being decidedly more intimate than any between them and other types. The remaining Oriental series also form a closely inter-related sub-group. It was surprising to find, as Dr von Bonin has observed, that the Tagals from the Philippine

* The reduced coefficients in Table IV may be compared with that of $8{\cdot}88 \pm {\cdot}19$ between the male Farringdon Street and Whitechapel series which both represent the population of London in the 17th century.

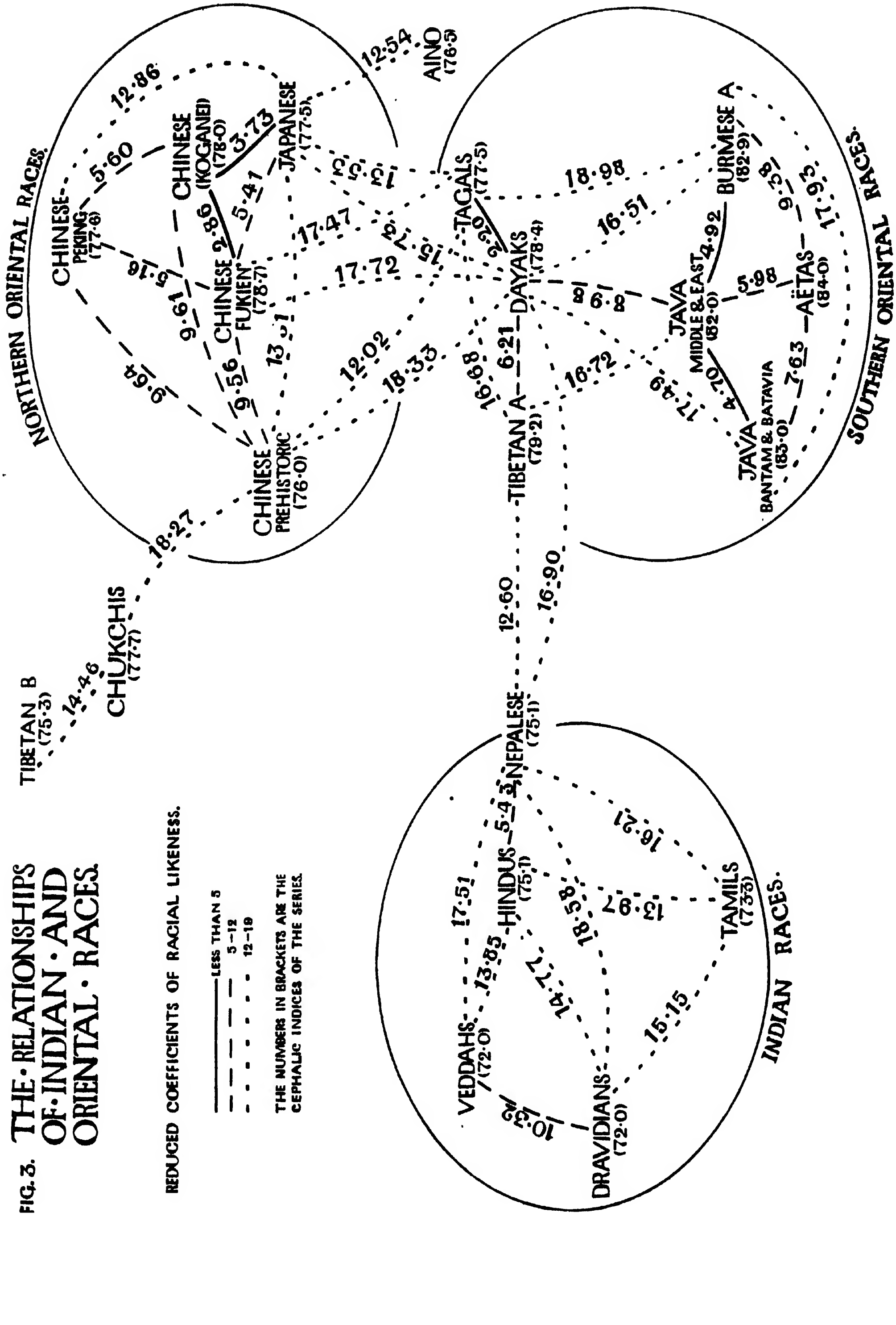


FIG. 3. THE · RELATIONSHIPS OF · INDIAN · AND ORIENTAL · RACES.

Islands and the Dayaks from Borneo are almost identical in type. The only reduced coefficients less than 19 between the two sub-groups are those connecting the Prehistoric Chinese, Fukien Chinese and Japanese with the Tagals and Dayaks. The close link between the last and the Tibetans of the *A* type is also an unexpected relation. The connection between the Dayaks and one series from Java, which is itself closely connected with another series from Java and the Burmese *A*, might have been anticipated; but the fact—previously noted by Dr von Bonin—that the Aëtas (a so-called negrito people from the Philippine Islands) are also intimately connected with the Javanese and Burmese is one of peculiar importance. The reduced coefficient between the Tagals and Aëtas is $26{\cdot}05 \pm {\cdot}65$, but the latter have links with other supposed non-negrito races of the Orient which are of precisely the same order as the lowest which can normally be found between one Asiatic series and another. This in itself provides sufficient justification for questioning the validity of the negrito hypothesis, and other evidence considered below raises the same doubts. It may be noted that there is, in general, no very close correspondence between the racial affinities of the series, as measured by these methods, and the geographical positions of the Oriental populations they represent, except in the case of the Chinese and Japanese groups. Migrations, of which some are known to have taken place in recent times, may well account for this condition. Contrasted with it is the extraordinary uniformity of the Chinese type. If more adequate material were available from remote parts of the country, there can be little doubt that distinct racial differences would be found, but there is already a clear suggestion that the greater part of the enormous population of China conforms more closely to a single racial type than do the present-day populations of several European countries*.

It has been suggested above that the coefficients of racial likeness make possible a division of the majority of the Asiatic series into three distinct groups, or four may be distinguished if the Oriental one is divided into two. Such a classification is, of course, only provisional and it is not unlikely that the apparent divisions between the groups will become less precise as more material becomes available. The arrangement arrived at is shown in Figs. 2 and 3, and it depends only on the evidence provided by the lowest reduced coefficients, none greater than 19 being considered in the diagrams. The experience derived from a similar comparison of European and other races has suggested that the most consistent results are to be obtained by considering only the closer degrees of affinity. The majority of the larger coefficients between the Asiatic series have yet to be presented, but before doing this it will be convenient to notice the order of the highest reduced values found between pairs of series belonging to the same group. In the case of the Northern Mongolian races (Table II) the highest coefficient is 51·07 between Telenghites and Mongols, but if the Telenghites are omitted the extreme is 25·01 between Buriats and Aleuts. The greatest divergence between two Indian series (Table III) is found in the case of

* Cf. "The Use of Biometric Methods applied to Craniology." *Biometrika*, Vol. XVIII. 1926, pp. 414—417. The coefficients of racial likeness are given in the above paper between a Northern and a Southern Chinese series compiled from various sources, but not used in the present paper, and the Fukien and Koganei's series. All are of a low order.

the Veddahs and Tamils having a reduced coefficient of 21·66. For the Oriental series the maximum value is 56·24 between the Chinese (Koganei) and Javanese (Bantam and Batavia): for the Japanese and Chinese alone the extreme is 13·51 in the case of the Japanese and Prehistoric Chinese, and for the other Oriental races it is 32·08 between the Tagals and Javanese (Bantam and Batavia). If the Oriental group is sub-divided in this way, and if the Telenghites are omitted from the Northern Mongolians, then the maximum reduced coefficient between two members of the same group is of the order 30: if these restrictions are not made the maximum is of the order 55.

We may turn now to Table V which gives the crude and reduced coefficients of racial likeness between the six Northern Mongolian and the five Indian races. The 30 reduced values range from 121·1 to 429·6, and the last appears to be the highest that has yet been found in comparisons between any pair of races in the world. Crude coefficients of racial likeness between all possible pairs of 41 European and Egyptian series have been published*, and when these 820 values are reduced the highest is found to be 174·0 between a British Neolithic and a Bavarian (Waischenfeld) series. Much greater divergences than this are evidently found in Asia. The greater reduced coefficients in Table V still, however, permit us to arrange the series in an orderly sequence. The Kalmucks, Telenghites, Buriats, Soyotes and Mongols all have their highest values with the Veddahs, their next highest with the Dravidians and so on in the order of the five Indian series given in the table. The Aleuts give the very similar order: Veddahs—Dravidians—Tamils—Hindus—Nepalese. In spite of the fact that all the Indian races are widely removed from all the Northern Mongolian races, there is thus clear evidence that the latter group, considered as a whole, resembles the Nepalese far more closely than it does the Veddah type. The linear arrangement given to the different Indian series by this means was not found to express their true relationships when they were compared directly. Reading the table in the other direction, it will be seen that there is less uniformity in the orders in which the six Northern Mongolian series are arranged by their reduced coefficients with the different Indian series. The lowest values are with the Kalmucks in three cases and with the Telenghites in the other two, while the Northern series furthest removed from the Indians are the Mongols and Aleuts. These arrangements such as can reasonably be accepted as indicating the true and somewhat complex relationships of the various types. The crude coefficients would suggest a different and a less consistent scheme of relationship which is far less likely to correspond to the actual racial links.

The 72 coefficients between the Northern Mongolian and Oriental series are given in Table VI. The range of the reduced values is from 43·2 (Burmese *A* and Telenghites) to 184·1 (Chinese Prehistoric and Aleuts), and there are 12 values less than 60. The closest connections here thus indicate a rather more intimate degree

* G. M. Morant: "A Preliminary Classification of European Races based on Cranial Measurements." *Biometrika*, Vol. xx^{B}. 1928, pp. 301—375. The reduced coefficients derived from the crude ones given in this paper have not been published.

TABLE V.

Coefficients of Racial Likeness between Northern Mongolian and Indian Races.*

		Crude Coefficients				
		Veddahs (33·8)	Dravidians (31·1)	Hindus (74·3)	Tamils (35·0)	Nepalese (45·9)
Kalmucks (33·0)	All characters ...	96·36 ± ·20 (23)	75·28 ± ·20 (22)	98·05 ± ·19 (24)	65·52 ± ·19 (25)	50·20 ± ·18 (29)
	Indices and Angles	104·12 ± ·34 (8)	97·46 ± ·36 (7)	98·70 ± ·34 (8)	87·84 ± ·32 (9)	53·52 ± ·29 (11)
Telenghites (55·3)	All characters ...	133·00 ± ·22 (19)	109·76 ± ·23 (17)	135·66 ± ·22 (19)	77·00 ± ·19 (24)	60·84 ± ·19 (26)
	Indices and Angles	241·81 ± ·43 (5)	247·33 ± ·48 (4)	250·87 ± ·43 (5)	150·04 ± ·36 (7)	110·31 ± ·34 (8)
Buriats (30·1)	All characters ...	108·32 ± ·22 (18)	88·88 ± ·23 (17)	99·85 ± ·22 (19)	60·40 ± ·20 (23)	55·24 ± ·21 (21)
	Indices and Angles	139·48 ± ·43 (5)	144·11 ± ·48 (4)	122·03 ± ·43 (5)	80·27 ± ·34 (8)	66·71 ± ·36 (7)
Soyotes (36·4)*	All characters ...	124·82 ± ·25 (14)	106·18 ± ·25 (14)	120·24 ± ·25 (14)	81·20 ± ·24 (16)	78·58 ± ·25 (14)
	Indices and Angles	158·77 ± ·48 (4)	141·38 ± ·48 (4)	161·88 ± ·48 (4)	127·49 ± ·43 (5)	119·03 ± ·48 (4)
Aleuts (28·3)	All characters ...	105·94 ± ·24 (16)	85·85 ± ·24 (16)	103·25 ± ·24 (16)	84·42 ± ·24 (16)	79·30 ± ·25 (14)
	Indices and Angles	139·15 ± ·43 (5)	121·70 ± ·43 (5)	131·15 ± ·43 (5)	137·45 ± ·43 (5)	127·69 ± 48 (4)
Mongols (109·4)*	All characters ...	238·60 ± ·28 (12)	185·47 ± ·28 (12)	286·77 ± ·28 (12)	154·72 ± ·25 (14)	147·51 ± ·28 (12)
	Indices and Angles	228·63 ± ·48 (4)	178·72 ± ·48 (4)	263·74 ± ·48 (4)	183·97 ± ·43 (5)	183·97 ± ·48 (4)
		Reduced Coefficients (all characters)				
Kalmucks ...		262·79 ± ·54	217·97 ± ·59	192·10 ± ·38	179·91 ± ·52	130·71 ± ·46
Telenghites ...		313·94 ± ·52	268·80 ± ·57	201·92 ± ·33	177·99 ± ·45	121·10 ± ·37
Buriats ...		320·42 ± ·67	272·23 ± ·71	215·35 ± ·47	186·51 ± ·61	154·09 ± ·58
Soyotes ...		335·53 ± ·69	315·14 ± ·76	235·25 ± ·50	227·42 ± ·67	190·84 ± ·62
Aleuts ...		337·47 ± ·76	299·33 ± ·83	255·50 ± ·59	269·69 ± ·76	230·33 ± ·74
Mongols ...		429·62 ± ·50	382·55 ± ·57	297·70 ± ·29	291·78 ± ·48	226·58 ± ·42

* See footnote to Table II.

of relationship than the most distant found between any pair of series belonging to the same group. The absence of any coefficients less than 40 suffices to make a trenchant division between the Northern Mongolian and Oriental groups. As a similar treatment of material from other parts of the world has shown, the most convincing and reliable classification is obtained by considering almost exclusively the closest degrees of relationship, and without attempting to reconcile the system built up in this way with one which might be derived from considering only the values of distant relationships. The reduced coefficient of 184·1 is greater than several found between the Tamils and Nepalese and the Northern Mongolian series (see Table V), but it is nevertheless true that the latter group, considered as a whole, resembles the Oriental far more closely than it does that of the Indian races. It might be expected that the Northern Mongolians would bear a closer resemblance to the Chinese and Japanese than to the other Oriental races, but this is not found to be true. The orders in which the reduced coefficients with the single Northern Mongolian series arrange the Oriental series are not closely similar—no two being exactly alike—but the Burmese, Japanese and Chinese (Koganei) tend to have the lowest values, while the highest are shown in most cases with the Prehistoric Chinese and Peking Chinese series. The different Chinese types are thus contrasted with one another in this way, and it is the southern type which resembles the Northern Mongolians more closely than the northern type. This inversion of the order we should have expected need not be emphasised, as all the coefficients concerned are high, and the conclusion that each Northern Mongolian race is equally related to all the Oriental races is, perhaps, the safest to accept in the present state of our knowledge. It is possible that some intermediate types may be found which will connect the two groups, but such are only likely to represent Manchuria and possibly Korea. It may be observed that the Telenghites and Kalmucks tend to be the Northern Mongolian series which most closely resemble the Oriental ones, while the Aleuts and Mongols are those furthest removed, but there is little uniformity in this matter.

Table VII gives the coefficients of racial likeness between the Indian and Oriental races. The reduced values range from 12·6 to 159·0 and there are 15 less than the lowest (43·2) found between a Northern Mongolian and Oriental series. In every case but one the Indian race with the lowest coefficient is the Nepalese, the exception being that the Tamils have a slightly, though not significantly, lower value with the Burmese A. It may be remembered that the Tamil skulls were collected in Singapore, though there can be little question as to their authenticity as all their closest relationships are with Indian series. The Hindus and Tamils appear to be almost equally removed from the Oriental populations, the Dravidians are more distant and the Veddahs most distant of all. Judging from all the coefficients, the Veddahs are rather further removed from the Oriental races (cf. Tables VI and VII) than are the Kalmucks, Telenghites or Buriats. By reading Table VII in a horizontal direction it can be seen that the Tibetan A, Dayak and Tagal series are the Oriental ones which most closely resemble the Indian, while the Chinese and Japanese group, on the one hand, and the Malayan and Aëtas, on

the other, are distantly and almost equally removed. The Indian and Oriental races are thus connected up by way of Nepal and Tibet, while there is no suggestion, as far as can be seen from this material, of a linkage south of the Himalayas. The division between the two groups (see Fig. 3) appears to be justified by the fact that the Tibetan *A* skulls were collected from the south of the country which is conterminous with Nepal, while the Nepalese have a decidedly more intimate connection with an Indian race and the Tibetans with an eastern one. It is quite possible, however, that this division would be found less marked if more abundant material were available.

A provisional classification into three groups is reached by these means. Every series has its closest connection, and in most cases all its closest connections, with other series in the same group. Importance is only attached to the closest degrees of relationship and there are numerous examples of two series belonging to the same group being further removed from one another than one, or both, of them is from series belonging to other groups. We are really dealing, of course, with a continuous system, and a diagram such as Fig. 3 illustrates the true state of affairs far better than any system of grouping can.

In the comparisons between the various Northern Mongolian races every series has one, or more, reduced coefficients less than 16, and if the Telenghites are omitted each series has one or more reduced coefficients less than 8 (Table II). The same limit for the Indian series (Table III) is 14 and for the Oriental (Table IV) 10. We may now consider the relationships of four other types which are supposed specialised since they cannot be placed in any one of the three groups of races distinguished. Crude and reduced coefficients of racial likeness are given in Table VIII. The Aino have one fairly close connection with the Japanese, the reduced value being 12·54. It may be suggested that this is sufficiently low to warrant the inclusion of the Aino in the Oriental group, but this seems inadvisable as the series has no other reduced coefficient less than 24 with members of that group. As has been suggested before*, the Japanese appear to have resulted from a cross between the Southern Chinese and the primitive island race, being now more closely related to the mainland type. If this be so, then the Aino may have had a very different origin from the bulk of the Oriental races. It may be noticed that they resemble the Tagals more closely than they do any Chinese race.

It is not surprising to find that the Chukchis from the extreme north-east of the continent are of another aberrant type. They are widely removed from the other Siberian races, and from the Aleuts, and the only reduced coefficient less than 20 with the series so far considered is the value of 18·27 with the Prehistoric Chinese. The three other Chinese series are almost as close, but it is suggestive that the prehistoric one should be distinguished from them in this way.

The Tibetan *B* skulls, from the eastern province of Khams, are known to represent a different race from those of the *A* type which come from the south-west of Tibet. The former series comprises 15 male specimens and the sample is hence

* *Biometrika*, Vol. XVI. 1924, pp. 61—62.

too small to lead to any reliable results. The few coefficients in Table VIII are suggestive, however. The only reduced value less than 20 is that of 14·46 with the Chukchis. All the Chinese types are further removed, but the closest connection here is with the Prehistoric series, and the Tibetan *A* type is more distant still.

The relationships of 25 Asiatic series (excluding the Tibetan *B*) have been considered above and the distribution of the lowest reduced coefficient found for each series ranges from a negative value (Kalmucks and Soyotes) to 15·55 (Kalmucks and Telenghites). The lowest with the Andamanese is that of 41·79 with the Tibetans *A*. There is, in general, a fairly close correspondence between the crude coefficient of racial likeness for all characters and for indices and angles alone, though a few marked differences may be found. Ignoring cases for which the second coefficient is based on fewer than 5 characters, there are found to be 5 examples in Tables II—VII of the value for all characters being more than 3 times the other. These are:—Javanese: Bantam and Batavia with Javanese: Middle and East (4·2 times), Dayaks with Tagals (5·5), Hindus with Nepalese (6·0); Chinese: Koganei with Chinese: Peking (11·1) and Tagals with Nepalese (4·1). The first three of these pairs are almost as closely related as any pairs of Asiatic series (see Fig. 3), the two Chinese types are also similar to one another and the Tagals and Nepalese are not widely removed. The same discordance is found between the Aino and Dayaks (3·0) and Aino and Nepalese (5·8), and a comparison of the means shows at once that it is the large size of the Aino skull which distinguishes it in these cases more than the shapes of its parts. Examining the Andamanese coefficients in the same way (Table VIII) it is found that the value for all characters is greatly in excess of the other in the comparisons with Aëtas (3·1), Chukchis (3·1); Javanese: Bantam and Batavia (3·7); Chinese: Fukien (4·0), Dayaks (4·0), Aino (4·5); Javanese: Middle and East (10·4) and Burmese *A* (15·0). But it is the small size of the Andamanese skull which is its most distinguishing characteristic. The shape is very similar to those of the neighbouring Burmese and Javanese. The close resemblance between the shapes in these cases cannot be supposed fortuitous and it seems to point clearly to the fact that the Andamanese came originally from Java, or from the neighbouring mainland, and that they have degenerated since. This theory cannot be reconciled with the negrito hypothesis however. It may be noted that of the four most specialised Asiatic types with which we are dealing two are found on islands while the Chukchis and Tibetans *B* represent peculiarly segregated races inhabiting regions which are among the most difficult to reach in Asia.

(4) *Comparison of Single Measurements.* When making comparisons between a considerable number of series, and at the same time making use of a considerable number of characters, as in the present case, the anthropometrician is accustomed to find that different measurements suggest incompatible schemes of relationship and there may be no sufficient reason why one of these should be trusted more than another. Use of the method of the coefficient of racial likeness frees him from this dilemma. It may be asked whether the classification suggested above can be

confirmed by a comparison of single measurements. The question whether a difference between two means is significant or not can be answered at once from the value of α between them found in computing the coefficients. The 31 characters compared in this way need be the only ones considered now. An α is supposed to indicate a significant difference if it is greater than 10, and its value, when greater than this limit, will indicate the degree of significance.

It will be sufficient to consider at the moment the evidence afforded by a comparison of single measurements within the group of 12 Oriental races which has been differentiated from other groups. It is found in this case, as is usual, that there are enormous differences between the characters in the capability of each to distinguish the racial types. Only 9 out of the 66 possible comparisons can be made in the case of the capacity and no one of these is significant. The foraminal index is available for all the series, but no α is found greater than 10. When it is found, in the case of a particular measurement, that less than 40 per cent. of the possible comparisons indicate a significant difference, it may be assumed that the measurement in question will be of little value either in suggesting a new arrangement of the series, or in confirming the one provided by the coefficients of racial likeness. In the present case, 21 of the 31 characters need not be considered in detail for this reason. These are, in addition to the two mentioned above, and in order of the percentage of α's greater than 10: fml (1·5), fmb (1·5), G_1 (1·8), 100 G_2/G_1 (2·2), $A\angle$ (10·7), J (13·6), $N\angle$ (14·3), G_2 (19·6), O_2 (21·8), 100 O_2/O_1 (22·6), B' (22·7), LB (24·2), B (28·8), U (30·3), $P\angle$ (33·3), 100 $G'H/GB$ (33·3), O_1 (34·0), 100 H'/L (35·9) and H' (35·9). No α's for these characters exceed 80. The means for the remaining ten are given in Table IX. For the cephalic and nasal indices and for the calvarial length more than half of the possible comparisons are found to indicate significant differences. In the upper part of the table the Southern Oriental series (see Fig. 3) are arranged in order of their cephalic indices and below them are the Chinese and Japanese in the same order. In spite of the clear distinction which was made between the two sub-groups, these distributions for the most variable single character are found to be overlapping. This condition is also found for all the other measurements except the nasal height and index—for which several means are missing—and for the upper facial index (100 $G'H/GB$). It is evident that the greater number of the clearly significant differences, and all the largest values of α, are found between comparisons of series not in the same sub-group. A direct comparison of the means might have suggested the division between the Chinese and Japanese, on the one hand, and the remaining Oriental series, on the other, but it would certainly not have led to any consistent scheme of relationship within each of these divisions such as those furnished by the lowest coefficients of racial likeness.

It may be asked whether the same characters are also those most capable of distinguishing the races within the other two groups. For the Northern Mongolian series the measurements showing 40 per cent., or more, of significant differences are: NH' (46·7 per cent.), 100 B/L (40·0), L (40·0), $G'H$ (40·0) and

TABLE IX.

Characters most capable of distinguishing Oriental series: means and α's.*

	100 B/L	100 NB/NH'	L	NH'	$G'H$	NB	S	$Oc. I.$	Q'	100 B/H'
Aëtas	84·0	54·1	171·0	49·5	69·6	26·8	360·0	65·0	323·7	105·7
Javanese: Bantam and Batavia	83·0	53·5	169·9	49·3	69·8	26·3	353·3	66·2	313·9	104·9
Burmese A	82·9	—	173·5	—	71·4	28·1	363·7	62·8	325·8	105·7
Javanese: Middle and East ...	82·0	53·1	173·7	50·6	70·9	26·8	360·4	65·4	318·4	105·1
Tibetans A	79·2	—	175·7	—	68·7	25·7	361·2	61·8	306·5	106·1
Dayaks	78·4	54·5	176·6	50·2	69·7	27·2	365·0	63·2	312·6	102·9
Tagals	77·5	55·5	179·2	50·0	70·0	27·7	375·0	62·6	317·5	101·2
Chinese: Fukien	78·7	48·1	179·9	52·6	73·8	25·2	377·0	61·8	322·0	102·3
Chinese (Koganei)	78·0	—	180·1	—	75·2	25·0	373·2	—	325·2	—
Chinese: Peking	77·6	—	178·5	—	75·3	25·0	370·0	61·1	317·0	100·5
Japanese	77·5	48·8	180·5	52·0	71·3	25·3	370·5	—	320·6	—
Chinese: Prehistoric	76·0	—	180·3	—	75·2	25·8	371·9	61·9	312·3	100·8
No. of α's	66	56†	66	57†	66	66	66	45	66	64‡
Per cent. of α's > 10	57·6	55·4	51·5	49·1	48·5	48·5	48·5	42·2	40·9	40·6

* All the means in this table are based on 23 or more male skulls.

† Including some comparisons between Frankfurt nasal heights (NH) or indices.

‡ Including some comparisons between 100 B/H''s.

O_1 (40·0). The measurements most capable of differentiating the Indian races from one another are: 100 O_2/O_1 (75·0), Q' (60·0), $G'H$ (60·0), NH' (42·9), 100 B/L (40·0), 100 H'/L (40·0), B' (40·0), and fml (40·0). There is little agreement between these two lists and that in Table IX, although the cephalic index and nasal and upper facial ($G'H$) heights are common to all three of them.

While no detailed analysis along these lines is likely to prove profitable, it will be convenient to grade the characters roughly according to their capability of differentiating Asiatic types. A comparison of the percentages of significant α's suggests the following arrangement, the groups referred to being the three dealt with in Tables II—VII above:

(*a*) Characters showing numerous marked intra- and inter-group differences throughout—100 B/L, $G'H$ and NH'.

(*b*) Characters showing numerous marked inter-group differences throughout and some marked intra-group differences—100 H'/L, 100 B/H', 100 NB/NH' and NB.

(*c*) Characters showing numerous marked inter-group differences throughout and hardly any marked intra-group differences—B, J, U, H' and O_2.

(*d*) Characters showing some marked differences in intra- or inter-group comparisons—L, B', Q', $Oc. I.$, 100 O_2/O_1, fml, S, 100 $G'H/GB$, fmb, $P\angle$, O_1 and $A\angle$.

(*e*) Characters showing few significant differences in any comparisons—$N\angle$, LB, 100 fmb/fml, C, G_1, G_2 and 100 G_2/G_1.

There are more measurements in division (*d*) than in any other and a comparison of these does not lead to any consistent arrangement of the types. Those in the first three divisions are the only ones likely to make any clear distinctions between the different groups. The ranges for these are given in Table X. If the Chinese and Japanese series are kept separate from the others in the Oriental group, as in this table, then there is no single measurement which makes absolute distinctions between all four groups. If all the Oriental series are grouped together, then B, J and $100\,B/H'$ are the only characters which make absolute distinctions between the three groups. It is clear that the cephalic index considered alone may be a

TABLE X.

Ranges of Mean Measurements for different Groups of Asiatic Races (*cf. Figs.* 2 *and* 3)*.

Races	$100\,B/L$	B	J	$100\,H'/L$
Northern Mongolian	81·4 – 86·4 (6)	148·7 – 151·5 (6)	139·8 – 144·0 (6)	70·0 – 73·7 (6)
Chinese and Japanese	76·0 – 78·7 (5)	138·2 – 140·9 (5)	132·2 – 135·6 (5)	76·0 – 77·0 (3)
Other Oriental ...	77·5 – 84·0 (7)	138·2 – 143·7 (7)	131·0 – 134·7 (7)	74·7 – 79·4 (7)
Indian	72·0 – 75·1 (5)	128·5 – 132·6 (5)	124·3 – 127·8 (5)	73·4 – 76·2 (5)

Races	$100\,B/H'$	H'	$G'H$	NH'
Northern Mongolian	113·6 – 118·0 (6)	127·6 – 131·1 (6)	71·3 – 77·6 (6)	52·4 – 56·6 (6)
Chinese and Japanese	100·5 – 102·3 (3)	137·0 – 137·8 (3)	71·3 – 75·3 (5)	52·0 – 52·6 (2)
Other Oriental ...	101·2 – 106·1 (7)	130·9 – 137·2 (7)	68·7 – 71·4 (7)	49·3 – 50·6 (5)
Indian	96·1 – 99·6 (5)	132·4 – 136·3 (5)	61·3 – 67·9 (5)	45·0 – 49·4 (4)

Races	O_2	NB	$100\,NB/NH'$	U
Northern Mongolian	34·1 – 36·2 (6)	25·5 – 27·5 (6)	48·4 – 51·4 (6)	516·1 – 524·9 (3)
Chinese and Japanese	33·8 – 35·5 (4)	25·0 – 25·8 (5)	48·1 – 48·8 (2)	502·2 – 511·6 (5)
Other Oriental ...	33·3 – 35·0 (7)	25·7 – 28·1 (7)	53·1 – 55·5 (5)	491·8 – 508·3 (7)
Indian	31·2 – 33·3 (5)	24·2 – 25·7 (5)	50·6 – 52·9 (4)	489·0 – 498·0 (4)

* All means used in compiling this table are based on 16 or more crania.

most unreliable guide to racial affinity. The ranges of the Northern Mongolian and Southern Oriental groups (excluding the Chinese and Japanese) overlap for this character, but the ranges for B, J, $100\,H'/L$, $100\,B/H'$, NH', $100\,NB/NH$, and U are discrete and quite widely separated.

(5) *Conclusions.* The classification of Asiatic races presented in this paper has been reached by using purely quantitative methods. In its broad outline it agrees with most other classifications which have been suggested by anthropologists using less exact methods. There are 26 cranial series for which adequate measurements

have been provided and a complete comparison was made between these by the method of the coefficient of racial likeness. The classification arrived at, shown in Figs. 2 and 3, was obtained by considering only the lowest orders of reduced coefficients. It has been found from the inter-comparisons of other groups of allied types that the most consistent results are always reached in this way. The Asiatic series can be divided into a number of distinct groups. The first of these (see Fig. 2) includes all from Mongolia and Siberia, with the exception of the Chukchi series from the extreme north-east of the continent, and one of Aleutian crania is also included. There appears to be a remarkable racial uniformity among a number of peoples dispersed over an enormous area stretching from Astrakhan to the Aleutian Islands. The series belonging to this group resemble one another closely and they are all markedly dissimilar from any other series available. The uninhabited desert of Gobi and the mountains to the north of Tibet are known to have shut off these Northern Mongolian peoples from their nearest neighbours to the south, and the craniological evidence suggests that there has been little, if any, admixture with the Northern Chinese. A second group of closely allied types is furnished by all the Chinese series available, including the Prehistoric series, and the Japanese. These show closer affinities to one another than to any races outside China and Japan, but no hard and fast line can be drawn between them and some Southern Oriental racial types. The Southern (Fukien) and Prehistoric Chinese and the Japanese are linked to the Tagals of the Philippine Islands and to the Dayaks of Borneo. The last two form a Southern Oriental group together with the Burmese, Javanese, Aëtas—a so-called negrito people from the Philippine Islands—and the Tibetans of the *A* type who inhabit the south-west of their country. A fourth group is made up by all the Indian series available, and no sharp division can be made between the primitive Veddahs and Dravidians, on the one hand, and the Hindus and Nepalese, on the other. The contiguous Nepalese and Tibetan *A* peoples are racially connected, though by a bond which is far less intimate than that between the Nepalese and Hindus, or that between the Tibetans and Dayaks. In this way all the Indian and Oriental types can be connected up to form, as it were, a continuous system with no wide gaps at any point, but with three centres round which there is a closer clustering of the types. There are three of the 26 series which cannot be assigned to any of these groups. The Andamanese conform to a type which is peculiar on account of its small size. But its shape is very similar to those of the Burmese and Javanese and it is reasonable to suppose that this indicates a true relationship, and that the stock degenerated after reaching the islands and there becoming isolated. An Aino series has only one close connection which is with the Japanese. The latter thus occupies an intermediate position between the Southern Chinese, to which it is far more closely related, and the Aino, and this is possibly an example of racial crossing. The third peculiar type is found not in an island but in the inaccessible region lying at the extreme north-east of the continent. The Chukchis are markedly dissimilar to all the series of the Northern Mongolian group and they resemble most closely that of the Prehistoric Chinese. The fact that there are no connections with any of the Modern Chinese

types is noteworthy. The Tibetan *B* type is presumed to represent the population of the province of Khams in Eastern Tibet, and this is another inaccessible region on account of its mountainous character. There are only 15 male crania available and coefficients were calculated solely with the series which obviously resemble the Tibetan *B* series most closely. The only connection found of the order we are now considering is one with the Chukchis.

This classification appears to us to be a reasonable one in every way. There is clearly a close association between the geographical positions of the peoples compared and the degrees of racial affinity between them as measured by the reduced coefficients of racial likeness. The segregation of the Northern Mongolian types and of the isolated Andamanese, Aino, Chukchis and Tibetans of the *B* type was to be expected owing to the peculiar nature of the regions they occupy. The connection of the Indian peoples with those of the Orient through the Nepalese and Southern Tibetans, and the connection of the Northern Chinese with the races of Java and Burma through the Southern Chinese, Tagals and Dayaks, are in close accordance with geographical considerations and they are clearly not accidental. The arrangement we have been led to by using purely quantitative methods has some unexpected features and the most striking of these is the close association of the Aëtas, who are generally classed as a negrito people having an entirely different origin from the non-negrito peoples of the Orient, with the Burmese and Javanese. We have also supposed that the Andamanese, who are also styled negrito, are closely allied to these two. If our deductions are correct then, as Dr von Bonin has suggested, the negrito hypothesis must be considered an entirely fallacious one in so far as it has been applied to Asiatic races. The evidence of the cranium is more likely to be a reliable guide in these matters than are characters such as stature and integumentary colours. The classification presented in this paper appears to be a suggestive one. The addition of further series may make it necessary to modify the details of the picture to a considerable extent, but we have every reason to hope that a reliable conception of the ethnic relationships and history of this and other groups of races will ultimately be reached by using statistical methods.

A STUDY OF THE CRANIA IN THE VAULTED AMBULATORY OF SAINT LEONARD'S CHURCH, HYTHE.

By BRENDA N. STOESSIGER, M.Sc. and G. M. MORANT, D.Sc.

(1) *Introduction.* The collection of skeletons at Hythe, representing the population of a single English town, is almost the largest one which can be examined to-day. Several anthropological descriptions of this material have been given at different times, but the only one which is of any permanent value whatever is that of Professor F. G. Parsons, published in 1908. Only a few measurements of 590 crania are given in that study and they are clearly insufficient to lead to any reliable conclusions regarding the racial affinities of the inhabitants of Hythe in past times. The main object of the present paper is to present detailed measurements of a sample of 199 of the specimens selected from the total of about 1500. We are greatly indebted to the Rev. C. W. Chastel de Boinville, Vicar of Hythe, for granting us permission to undertake this work and for aiding us in other ways. We have also attempted to give a more comprehensive and accurate history of the human remains and a fuller account of the evidence bearing on the ethnic history of the population they represent than any previously published.

The technique used in taking the measurements, preparing the type contours and estimating the affinities of the racial types compared was precisely the same as that used in the study of the Spitalfields Crania*, and the reader is referred to that paper for particulars regarding these matters and for references to many of the sources of the comparative material. Some use was also made in the Spitalfields paper of our measurements of the Hythe skulls which are here published in detail.

(2) *The History of the Town and People of Hythe.* There are two related historical questions relevant to the present inquiry which have to be considered. The first concerns the Church of St Leonard and its human remains directly, and the second the nature and racial constitution of the population of Hythe at different periods. Unfortunately the town has not yet found its historian and no authoritative ruling has been given on many of the points we have had to consider. The compilation of the material on which anything approaching a comprehensive history might be based has, indeed, yet to be made. It is clear that the characters of the early population of Hythe cannot be estimated without some reference to the surrounding district, and the map (Fig. 1) shows the region with which we shall be principally concerned. Almost due west of St Leonard's Church, and at a distance of 2 miles, is West Hythe, now a hamlet with the ruins of the Church of

* G. M. Morant, with assistance from M. F. Hoadley: "A Study of the recently excavated Spitalfields Crania." *Biometrika*, Vol. xxiii. 1931, pp. 191—248.

St Mary the Virgin. Less than a mile beyond is Stutfall Castle with Lympne Castle and village to the north. Passing eastward from Hythe we come first to Seabrook, where a small stream which was once considerably larger now flows into the Royal Military Canal, then to Shorncliffe Camp, formed during the Napoleonic wars, and finally to Sandgate, $2\frac{3}{4}$ miles from our point of departure, where there is a castle built in Henry VIII's reign. The line from Stutfall Castle to Sandgate rises above 100 ft. in places, though the old town of Hythe, built on the side of the hill, was nearly all below the present 50 ft. contour line. This worn down undulating ground was

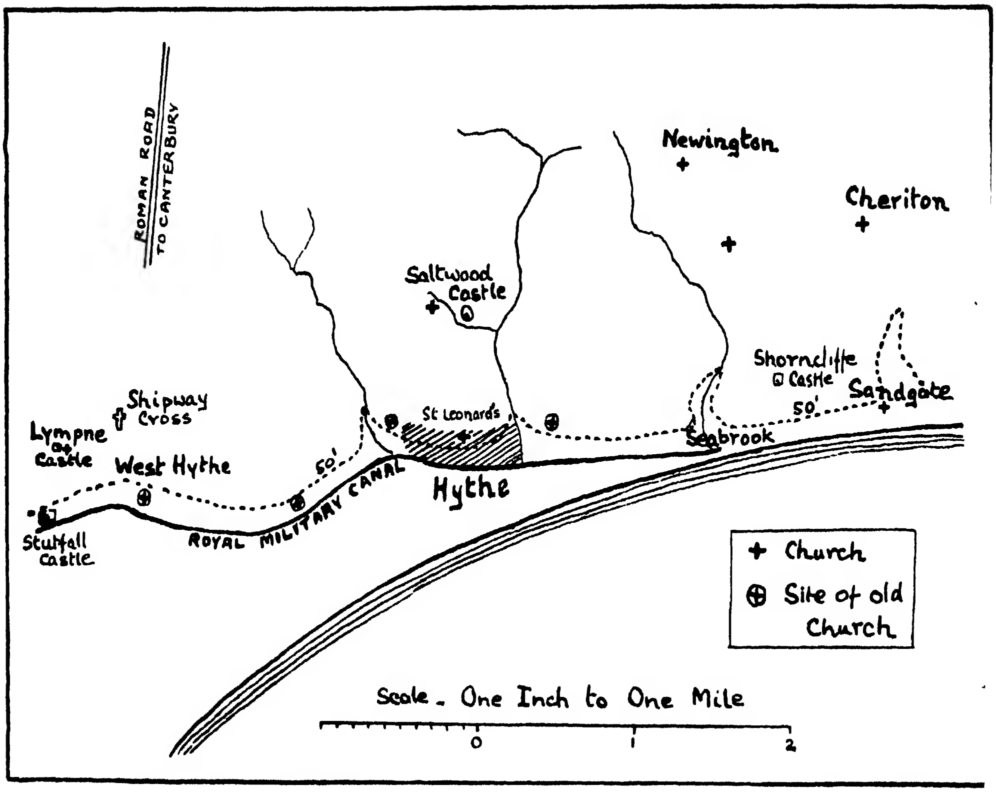


Fig. 1.

formerly the escarpment of Lower Greensand deposits and it is known to have been higher, and to have had a steeper slope in parts, within historical times. There are records of landslides—called earthquakes by some of the earlier writers—which have affected it and Stutfall Castle was injured by a subsidence at some unknown date. The Greensand ridge comes down to the sea at Sandgate, but to the west it is separated from the present-day coast-line by a plain which begins at Seabrook, increases in width to the west and finally merges into Romney Marsh. The northern edge of this area, which is nowhere more than 20 ft. above sea-level, is marked roughly by the line of the Royal Military Canal. The old town of Hythe was situated half a mile, and Stutfall Castle is $1\frac{1}{2}$ miles from the present

shore. The coastal plain, like the marshes to the west of it, is marked on geological maps as alluvium, but, although few borings have been taken, there can be no doubt that it was formed by the combined action of river and sea. Its outline is known to have been modified to a large extent during historical times owing to both human and natural agencies, and these changes have had a profound effect on the population of the district at different periods. Several incompatible explanations of the process which resulted in the formation of the eastern end of Romney Marsh have been suggested*. The Rhee wall from Appledore to New Romney was constructed in pre-Roman or Roman times, coins and other antiquities of the occupation having been found in many parts of Romney Marsh which it protected. An early theory, championed by Holloway†, was that the principal outflow of the marsh was the river Limen at this time, and it is supposed to have had a wide estuary immediately below the hills on which Lympne stands, while there was a Roman port at that place. The sea is thought to have receded from Lympne at the beginning of the seventh century and this led to the rise of West Hythe, at which place the river still debouched. Holloway supposes that by the end of the eighth century the Limen had become divided into two branches, one still flowing out at West Hythe and the other—the Rother—at Romney. It is conjectured that the bed of the Limen dried up some time during the next two or three hundred years, but traces of it were said to be still visible when Holloway wrote. According to another early theory, championed and elaborated by Lewin, there was no river or branch of the sea below Lympne Hill in Roman times, for water there would have been at a higher level than the marsh. He remarks that: "Roman remains are scattered over the whole of Romney Marsh, and may be found in every field that is ploughed‡." A great shingle spit is supposed to have reached all along the border of this area from Lydd to Sandgate with one break between Lydd and Romney and another at Hythe. Behind the latter there was a narrow gut extending from West Hythe to Shorncliffe. The harbour formed in this way was kept scoured by three streams, of which one was the Seabrook, which

* The question is dealt with in the following sources: William Somner, *A Treatise of the Roman Ports and Forts in Kent*, 1693, pp. 41—61. William Holloway, *The History of Romney Marsh from its earliest Formation to 1837*, 1849. Charles Roach Smith, *Report on Excavations made on the Site of the Roman Castrum at Lymne, in Kent, in 1850*, 1852, pp. 39—45. Thomas Lewin, *The Invasion of Britain by Julius Caesar*, 1859; *The Invasion of Britain by Julius Caesar with Replies to the Remarks of the Astronomer Royal and of the late Camden Professor of Ancient History at Oxford*, 1862; "On the Position of the Portus Lemanis of the Romans," *Archaeologia*, Vol. XL. 1866, pp. 361—374. More recently Dr T. Rice Holmes has discussed at considerable length the numerous theories which have been held regarding the changes Romney Marsh has undergone in the past 2000 years (*Ancient Britain and the Invasions of Julius Caesar*, 1907, pp. 532—552, 622—625 and 640—641). He gives numerous references and the geological evidence is considered. It is curious that the most complete inquiries regarding these matters have been made with the object of substantiating some theory or other—and often one which is really of minor historical importance—relating to the Roman occupation. Lewin attempted to prove that Julius Caesar landed near Hythe, and Rice Holmes only deals with the district in order to refute this view. Holloway endeavoured to show that Anderida was at Newenden, beyond the west end of the marsh, and not at Pevensey.

† *Op. cit.*, pp. 14—23, 47—49, 52—56, 62, 65 and 102.

‡ *Loc. cit.*, 1866, p. 366.

flowed into it from the north. These were unable to perform the work effectively and there was a gradual diminution in the size of the Hythe Haven*. The theory that there was a Roman port below Stutfall Castle is said to be wholly untenable. The castrum there was built to defend the marsh and the port was not there, or at West Hythe, but at Hythe. After an examination of a great deal of evidence, part of which had not been accessible to Lewin, Rice Holmes came to the following conclusions: "first, that the Rother did not, in the time of Caesar, enter the sea at Lympne...; secondly, that the marsh was then closed at West Hythe Oaks (half way between Hythe and West Hythe), and therefore that there was no harbour at Lympne; thirdly, that the Rhee Wall had not then been built...; fourthly, that the Portus Lemanis was a pool harbour extending from West Hythe to a point nearly opposite Shorncliffe...†." There is thus a substantial agreement between the views of Lewin and Rice Holmes with regard to these matters. Rather different conclusions have been reached by a number of other writers. Everyone admits that a branch of the sea came close to the hills along part of the line from Hythe to Stutfall Castle in Roman times, and it has been generally held that this estuary reached West Hythe and that the port was there, although there may have been no river running along the north of the marsh.

There is a certain amount of historical and archaeological evidence relating to Roman times which bears directly on this district. Stutfall Castle is the Saxon name of a building which is entirely Roman. The following sources have been supposed to refer to this castrum, or to the immediate neighbourhood‡: the *Geography* of Ptolemy refers to the "New Port" or "New Haven" which has been identified with Lympne, though this view has been contested; the *Antonine Itinerary*, assigned to the second or third century, gives the distance from London to the "Portus Lemanis" and between other towns on the way, treating it as one of the three Kentish ports, the other two being Richborough and Dover; the *Notitia Dignitatum*, compiled at the beginning of the fifth century, refers to "Lemanis" as the place where an officer of a detachment of the *Turnacenses* held a garrison under the command of the Count of the Saxon Shore; the *Peutinger Table*, assigned to the latter part of the fourth century, mentions the same place with the symbol of a gateway between towers signifying a fortified city or port;

* Some of the sixteenth and seventeenth century maps of Kent show the town of Hythe with St Leonard's Church above, a few houses below it, a stream running down on either side of these buildings and the two meeting in a pool, with one or two islands in it, which was situated between the town and the sea. According to Lewin (*op. cit.*, 1862, p. lviii) the position of the mouth of this pool could be clearly seen when he wrote, the hollow having been filled in a few years before.

† *Op. cit.*, p. 552.

‡ In addition to the sources already cited, the following are the more important ones dealing with the district in Roman times: Charles Roach Smith, *The Antiquities of Richborough, Reculver and Lymne, in Kent*," 1860. William Henry Black, "On the Identification of the Roman Portus Lemanis," *Archaeologia*, Vol. XL. 1866, pp. 375—380. George E. Fox, "The Roman Coast Fortresses of Kent," *The Archaeological Journal*, Vol. LIII. 1896, pp. 352—375. R. F. Jessup, *The Archaeology of Kent*," 1930. Francis Hobson Appach, *Caius Julius Caesar's British Expeditions from Boulogne to the Bay of Apuldore, and the subsequent Formation geologically of Romney Marsh*," 1868. Appach attempted to prove that Romney Marsh was not formed until the middle of the fifth century A.D.

the anonymous Geographer of Ravenna mentions "Lemanis." All later writers agree in identifying the fort of Lemanis with Stutfall Castle. The *Portus Lemanis* is only mentioned clearly in the *Antonine Itinerary*. Somner thought that it was at Romney*, but all later authorities reject this view and place it either below Stutfall Castle, or at Hythe, or somewhere between these two places. The theory that it was at West Hythe appears to be the most plausible. The Roman road from Canterbury, known as Stone Street, is still well defined for over ten miles and it runs straight towards Shipway Cross (see Fig. 1) and the modern hamlet of West Hythe. The first excavation of Stutfall Castle was undertaken by Roach Smith in 1850 and it was very incomplete owing to the lack of sufficient funds. The site is a large one covering more than ten acres and most of the walls above ground had been removed at an early date and some of the materials were used in building Lympne Church and Castle. Landslides have distorted the remains. The chief entrance was found to be on the eastern side. The only building of interest discovered inside the boundary walls was one which had been provided with hypocausts and two fire-places. It has been suggested that this was either an officer's house or the baths of the station. Tiles which had been previously used in constructing another building and an altar forming part of the foundations of a gateway bore inscriptions which are interpreted as *Classiarii Britannici*, or British marines. The altar was covered with barnacles, proving that it had been washed by the sea at one time. The age of the fortress could not be placed earlier than Constantine, owing to structural details, and it was hence one of the last Roman stations built along the south coast. It was also one of the largest and most important. At some date, probably before Constantine, a division of the British fleet must have been stationed in the vicinity, and later the large castrum was built and occupied at one time by Turnacensians (from Tournai). In 1894 further excavations on the site of the castle were carried out by Professor (later Sir Victor) Horsley†. The absence of the south wall had been taken to indicate that there was a river or sea defence on that side, but vestiges of the southern wall were found at this time, proving that the castrum was of the usual quadriform pattern. Numerous other Roman remains have been found in the locality with which we are concerned. Leland says in his account of Stutfall: "About this Castel, yn tyme of mind, were found Antiquites of mony of the Romaynes‡." The numerous remains found in the adjacent marsh have been referred to (p. 137 above) and there were potteries at Dymchurch less than three miles away. Roach Smith§ says that there was no record in 1850 of Roman remains having been found in the vicinity of the castrum, and its burial-place had not been discovered, but he refers to the discovery of a building of this period less than two miles away to the north-west. In attempting to trace the course of the Roman road from the villas he

* *Op. cit.*, pp. 37—62.

† The only account of these is a paragraph reporting a lecture in *The Athenaeum*, 22nd September, 1894, p. 394.

‡ Thomas Hearne's edition of *The Itinerary of John Leland the Antiquary*, Second edition, Vol. VII. 1744, p. 132.

§ *Op. cit.*, 1860, pp. 262—264.

discovered at Folkestone in 1924 to Lympne, S. E. Winbolt mentions a number of finds in the locality*; these are: "a coin of Marcus Aurelius found (1904) 18 in. down in Hill Crest Road, Hythe; Roman burials found in the quarry at the corner of Hill Crest Road and Castle Road (both of these at the top of the slope above North Road); Roman remains found in Harp Wood in 1874...and Roman remains in the glebe land of Lympne Vicarage." It is inferred that the old road went along the line of North Road immediately above St Leonard's Church. The corner of Hill Crest Road and Castle Road is less than 200 yards from the church. Harp Wood is nearly a mile away to the north-west†. There is another record of Roman remains having been found in Hythe. In his book published in 1862 Lewin says that: "in excavating for a drain at the east end of Hythe, we came to the foundations of a Roman building in the main-road, about two feet under the surface, and turned up at the same time a great quantity of broken Roman pottery‡." The main road runs below St Leonard's Church. This writer attempted to prove that Caesar landed at Hythe and that the subsequent battle with the Britons took place "in the field to the south and east" of the town. We are told that "on the triangular level there, human bones, and unquestionably of men slain in battle, are brought to light. They are exclusively the bones of grown men, buried only a few feet below the surface, and without any care, in all conceivable positions. I do not affirm that these are the remains of the Britons who fell in the conflict with Caesar, for they may be the bones of either Saxons or Danes who afterwards landed at the same place...§." Mr Elliott is cited as having supplied this interesting information‖. The land in question is said to be distinct from the great shingle bed and below high-water mark. It would appear to have been under the water of the gut which Lewin thought to be the port in Roman times. He recognises this and observes that it was certainly dry at low water. The evidence reviewed above seems to indicate conclusively that there was a considerable population at Hythe during part, at least, of the occupation. Stutfall Castle must have been an important station with a large garrison towards the end of that period, and it is

* *Roman Folkestone*, 1925, pp. 158—160.

† The site near Harp Wood is marked on the 6-inch Ordnance map.

‡ *Op. cit.*, 1862, p. cxxi.

§ *Ibid.*, pp. lxxiii—lxxiv.

‖ The accuracy of Lewin as a reporter or observer is certainly not above suspicion. On p. 92 of the second edition of his book he says that the sea flowed below Stutfall Castle in ancient times up to the very base of the hill, "as is proved incontestably by the fragments of ships and anchors which have been dug up...." Rice Holmes (*op. cit.*, p. 622) points out that on pp. lxviii and lxix of the same book it is denied that the sea was ever below Stutfall Castle in Roman times and the evidence of the fragments of ships and anchors is entirely ignored. Lewin's lack of consistency is, in fact, far more thoroughgoing than this, however. In his later paper in *Archaeologia* (pp. 364—365) we are told that, if the Portus Lemanis had been at the foot of Lympne Hill, "we should expect to find at least some vestiges, however faint, of the port itself. The ground there has been long under cultivation, but I have never heard or read (though I have often inquired) that any remnant of a pier or sunken vessel, or even any anchor or other part of a ship's tackle, was ever discovered in this part. Again, had the port existed here, the adjacent parts on the hill side must have been covered with wharves and warehouses and the dwellings of the seafaring population; but, with the exception of Stutfall itself, no signs of population here show themselves."

probable that in the earlier centuries, at least, there was one of the most important ports of the country situated somewhere between the castle and the modern town. Mercenaries from Flanders are known to have been quartered in the castrum at one time, but there is no other indication of the racial constitution of the population during this epoch.

On turning to the history of the district in Saxon times we are at once confronted with another problem which has aroused controversy among historians. Hasted gives the following account of the matter:

"During their contests, in the year 456, a bloody battle was fought near this place, between Folkestone and Hythe, between the Britons under K. Vortimer, and the Saxons, who were retreating hither before him, after the conflict he had with them on the banks of the Darent, in the western part of this country. According to some writers, this battle was not fought near Folkestone, but in Thanet; but as the Britons drove the Saxons, after the battle, into that island, the place of conflict could not be there. Nennius and others write, that it was fought in a field on the shore of the Gallic sea, where stood the *lapis tituli*, which Camden, Usher, and Baxter, caught by the sound of the name, take to be Stonar, in that island; but Somner, Gale and Stillingfleet, instead of that, read, in a correction of their own conjecture, *lapis populi*, or Folkestone. This place certainly suits best with the description of it, on the shore of the Gallic sea; and what adds strength to this, are the two vast heaps of sculls and human bones, piled up in two vaults under the churches of Folkestone and Hythe, which, from the quantity of them, could not but be from some battle; and, from their whiteness, appear to have been all bleached by lying for some time probably on the sea shore; and many of the sculls have deep cuts in them, as made by some heavy weapon. Probably those at Hythe were of the Britons, and those at Folkestone of the Saxons, who were pursued hither by them *."

Hasted was not the originator of this theory which is supposed to account for the existence of the bones at St Leonard's, though he does not acknowledge the fact. The authority of this most voluminous, though perhaps not most accurate, historian of Kent was very generally accepted without question for many years after he wrote and, as he offers only one explanation with regard to this matter, his tale—though really absurd—was often quoted as if it stood for the final verdict of posterity. We are asked to believe that a battle took place somewhere between Folkestone and Hythe and that there were some thousands on each side slain. Some time shortly afterwards, presumably, the corpses were sorted out, all the Britons being collected in one place and all the Saxons in another. They were then left until the bones had been bleached and at some time *at least 700 years* later the skeletons of the Britons were transported to Hythe and those of the Saxons to Folkestone, and it should be remembered that these two towns are seven miles apart. If any denial of such a preposterous theory is needed it is found in

* Edward Hasted: *The History and Topographical Survey of the County of Kent*, Vol. III. 1790, p. 378, footnote (*e*). A similar account is given on p. 420 of the same volume and on p. xxviii of Vol. I. 1778.

the fact that there are considerable numbers of women and children represented by the bones at Hythe. The only documentary evidence which could suggest that a battle was fought in the district at this time is a short passage in the *Historia Brittonum*, a composite work of which parts were written as late as the middle of the ninth century*: The *lapis tituli* on the shore of the Gallic sea is not identified with Folkestone by modern scholars. As far as we have been able to ascertain, the first writer to suggest that the bones at Hythe might possibly be associated with the fifth century battle was John Harris†. His conjectures regarding the matter are far more sober than those of Hasted. The origin of the remains is said to be unknown, but two guesses may be ventured on; the first is that they came from graveyards in the town and the second that they were "collected and piled up here on some eminent occasion." Of the two "eminent occasions" suggested one is the battle between the Britons and Saxons. The bones may be those of the soldiers of the two armies, "whose bodies fell herabouts and at Folkestone" and the existence of another ossuary there renders this supposition more probable‡. This supposition may be dismissed as being entirely untenable§.

The following inscription was hanging in the ambulatory passage of St Leonard's at the beginning of the nineteenth century:

"From an antient History of England brought down to 1658.

A.D. 853. The Danes landed on the coast of Kent, near the town of Hyta (now Hythe).... They were...at length defeated by Gustavus, the governor of Kent, who assembled the greatest part of the inhabitants, assisted by the army of Ethelwolf, then king of Britain, who met the invaders near Hyta, when the Danes...being overpowered fled to their vessels, then on the coast near the above town; but being closely pursued, they made a bold stand near the water, where the battle became general, and tradition reports that upwards of 30,000 fell in the conflict. After the battle, the Britons...returned to their homes, leaving the slain on the field of battle; where being exposed to the different changes of the weather, after a length of time the flesh rotted from the bones, which were at length collected and piled in heaps by the inhabitants, who in time removed them into a vault in one of the churches of Hyta. D. Thomson, A.D. 1797||."

* H. Munro Chadwick: *The Origin of the English Nation*, 1924, p. 36.

† *The History of Kent*, Vol. I. (the only volume printed) 1719, p. 152.

‡ Harris says that he had been told that in digging a grave at Folkestone Church a vault was found "where great quantities of bones, like these (at Hythe), were piled up." Mention of this vault was made by Thomas Philipott (*Villare Cantianum*, 1659, p. 96). Local interest in ossuaries was likely to be lively and there are several references to visitors who attempted to locate this one, and to see its contents, in the eighteenth and nineteenth centuries, but without success. S. J. Mackie (*A Descriptive and Historical Account of Folkestone and its Neighbourhood*, 1856, p. 105) says that the crypt containing bones was found "a few years since" under the north chancel.

§ In describing the final stages of the conquest of the Jutes in Kent, John Richard Green (*The Making of England*, 1881, pp. 39—40) supposes that the last years of Hengest were occupied in reducing the fortresses on the southern coast and that of Lympne is mentioned as being the last to fall. There appears to be not a shred of documentary or archaeological evidence, however, to show that Stutfall Castle was ever occupied, or defended, in the second half of the fifth century.

|| From a letter in *The Gentleman's Magazine*, Vol. 72, 1802, pp. 1001—1003.

The earliest reference to this inscription we have been able to find is one made by Charles Seymour in his survey of the county published in 1776*, so the one above was probably copied from an earlier original. Another copy with slightly different wording was made in 1812 and this was framed and may be seen hanging in the ambulatory to-day. The date of the battle is given as 843 and the signature "D. Thomson, A.D. 1797" is omitted†. Several other writers have referred to, or given transcripts of, this record and the dates 842 and even 143 have been given in error in some of these accounts. D. Thomson was probably the transcriber of the original account. We have not been able to identify the "antient History of England (or 'Britain' in the 1812 copy) brought down to 1658," or to find any author who has suggested that a battle with the Danes was fought at or near Hythe in the middle of the ninth century. The invaders were defeated by Ethelwulf, leading the West Saxons, at Aclear (Ockley) and it is just possible that this is the event referred to. The *Anglo-Saxon Chronicle* gives the years 851 and 853 in different MSS. and the former is now generally regarded as being the authentic one. We can only conjecture that the authority of a particularly imaginative ancient historian led to a wildly improbable theory when the bones were associated with this particular battle. The theory was generally accepted until recent years.

During the second Danish invasion, and after the death of Ethelwulf, a considerable force was landed at the mouth of the Limen. One writer has suggested that this was at Hythe, but his theory has not been generally accepted, even by those who believe that there was a river estuary below the town in post-Roman times. Two chronicles mention the *Portus Lemanis* in this connection, but the *Anglo-Saxon Chronicle* states that the landing in 893 was at the mouth of the Limen at Appledore‡. Two charters of 732 and 833 refer in almost identical terms to a piece of land bounded on the south by the Limen and having on the north the "Hudan Fleot."‡ The last has been identified with West Hythe, but this is merely conjectural. According to an early account which has been repeatedly copied, the manor of Hythe was given by King Alfred§ to the Priory of Christ Church, Canterbury, in 849, but this appears to be another spurious tale. This transference was made,

* *A New Topographical, Historical, and Commercial Survey of the Cities, Towns, and Villages of the County of Kent*, p. 477.

† S. J. Mackie tells us that the memorial—probably meaning a copy of the original—was "written in a fine hand by the favourite pupil of a local pedagogue" (*op. cit.*, p. 180). He also refers to "Roman and Saxon pottery and mediaeval coarse earthenware" found in restacking the pile of bones at Hythe and in his possession when he wrote. Thomas Wright (*Wanderings of an Antiquary*, 1854, p. 120) also refers to the pieces of pottery found, "some of which are of a very early character, and appear to me like fragments of Anglo-Saxon burial urns. Among them were some fragments of glazed mediaeval pottery of a later period—probably of the sixteenth century...."

‡ See references and discussions by Rice Holmes (*op. cit.*, pp. 539—542) and Holloway (*op. cit.*, pp. 18—20).

§ Alfred was born in 849. Kilburne (1659) appears to have been the first to mention this grant and it has been referred to—without references, as usual—by many later writers. Hueffer (*The Cinque Ports*, 1900, p. 191) gives the date as 889, which is more reasonable, but he gives no authority for the statement.

however, in 1036 by Halden, or Halfden, a Saxon thane*. There appears to be only one other undisputed reference to the town before the Conquest. Two MSS. of the *Anglo-Saxon Chronicle* give an account of the revolt of Earl Godwin in 1052. He sailed with Harold from the Isle of Wight and visited Pevensey, Romney, Hythe (Hide or Hythe), Folkestone, Dover and Sandwich and "even took all the ships that they found, which might be of any value, and hostages as they went, and then betook themselves to London...†." All the towns, except Hastings, which later became the Cinque Ports are mentioned here together with Pevensey and Folkestone and it is probable that all the more important Kentish ports of the time were visited.

Few Anglo-Saxon remains have been discovered in the neighbourhood of Hythe. A disused quarry to the north-west of the town is marked on the Ordnance map as the site of finds of this period in 1870, but no other record of the event appears to have been made‡. Burial places have been excavated at Folkestone, and near Lympne. Romney Marsh was probably uninhabited during this period, though the east and north of Kent were more densely populated by the invaders than any other part of England§. The origin of the town of Hythe, as of most of like antiquity, is obscure, but it is at least evident that it was a port of some importance before the Norman Conquest.

Further evidence of the early importance of the town is furnished by the *Domesday Book*; the following are the references to Hythe:

"Hugh de Montford holds, of the Archbishop, Salteode....To this Manor pertain two hundred and twenty-five Burgesses in the borough of Heda." ¶

"The same Archbishop holds Leminges in demesne....Thereto pertain six burgesses in Hede¶."

The aggregate value of the Borough of Hythe and Manor of Saltwood was estimated at sixteen pounds in the reign of Edward the Confessor and eight when it was transferred. At the time the survey was made, its total produce was said to be twenty-nine pounds six shillings and four-pence. It is generally recognised to-day that no reliable estimate of population can be deduced from the *Domesday Book*. Most of the particulars relating to the towns were probably only entered because of their bearing on the fiscal rights of the Crown. We find that Dover, Sandwich and Romney are mentioned as providing sea service, but Hythe and Hastings are not. In the account of Kent, reference is made only to 816 burgesses,

* Hasted: *op. cit.*, Vol. III. p. 141. Dugdale: *Monasticon Anglicanum*, 1682, Vol. I. p. 21. This is the earliest reference to Hythe in Saxon times mentioned by Hasted. The Rev. H. D. Dale, sometime Vicar of Hythe, says that it is the first undisputed reference to the town ("Notes on Hythe Church," *Archaeologia Cantiana*, Vol. XXX. 1914, p. 275).

† Thorpe's edition, 1861, Vol. II. p. 153.

‡ See the article on "Anglo-Saxon Remains" by Reginald Smith in *The Victoria History of the Counties of England, Kent*, Vol. I. 1908.

§ A map showing the distribution of Anglo-Saxon burial places for the whole of England has been given by E. Thurlow Leeds: *The Archaeology of the Anglo-Saxon Settlements*, 1913, p. 19.

¶ Rev. Lambert Blackwell Larking: *The Domesday Book of Kent. With Translation, Notes and Appendix*, 1867, pp. 103 and 105.

438 of these belonging to Canterbury, 231 to Hythe and 135 to Romney. But the county survey opens with a long description of Dover which does not detail a single burgess in the town, and there may have been similar omissions for other places. In spite of the meagre notice of Hythe in this record, it is reasonable to infer that the town was one of the largest and most important in Kent in the eleventh century.

Further evidence that Hythe was an important centre before the Conquest is furnished by the fact that it was one of the original Cinque Ports*, together with Sandwich, Dover, Romney and Hastings, and all these were amongst the most flourishing and populous English towns in mediaeval times. The origin of what was once the most important corporation in England is obscure owing to the loss of the earliest charters in the sixteenth and seventeenth centuries. The earliest extant is that of the sixth year of Edward I (1278) and it refers to earlier ones granted by Edward the Confessor, William I and other kings. The Ports grew in importance and they were most flourishing under Edward I, who fostered them. Disasters from which they never recovered began in Edward III's reign, but until the end of the fifteenth century the corporation had to furnish the Crown with nearly all the ships and men needed for the State. The primary cause of their ultimate decline was the silting up of their harbours, and Romney and Hythe became affected before Sandwich and Dover. The "Two Ancient Towns of Winchelsea and Rye" became members with the same status as the original five; and other subordinate members, or limbs, were attached at various times to the seven principal Ports. Most of these affiliated groups were made up by five, or more, small towns, but West Hythe was the only one ever attached to Hythe, and it was a non-corporate member within the municipal jurisdiction of the larger town. The services demanded from the Ports in return for the immunities and privileges they enjoyed are known for some years and they show little variation during the earlier centuries. In 1229 a total of 57 ships had to be provided, each manned by 21 men and a boy. This number of ships was assembled from the Ports at several later dates, but the crew in each had to be increased to 24 and later to 34. The following list, showing the proportions of the total fleet which were contributed by Hythe in different years, is compiled from figures given by Jeake and Burrows:

Year	Total number of ships provided by the Cinque Ports and their members	Number provided by Hythe
1229	57	5
1294	50	3
1300	30	4
1347	105 (?)	6
1360	57	5

* The following are the most important sources of information relating to the Cinque Ports: Samuel Jeake, *Charters of the Cinque Ports, Two Ancient Towns and their Members. Translated into English, with Annotations Historical and Critical thereupon.* This book was written in 1678 and published posthumously in 1728. It is still by far the most important source of first-hand information on the subject. Sir Nicholas Harris Nicolas, *History of the Navy from the Earliest Times to 1422*, 1847, 2 vols. Montagu Burrows, *Cinque Ports*, 1888. This is by far the most authoritative modern work on the Ports,

The total number in the fleet at the Siege of Calais in 1347 was 710 ships and 14,151 men, so the Ports were only able to supply a proportion of the war-time strength needed at this time. The Cinque Ports had supplied only a proportion, too, of the 200 English vessels which took part in the Battle of Sluys (1340). In the following year the shipping demanded from Hythe and Romney was not ready and their franchises were ignored for a time. In 1344 an exceptional demand was made and all the Ports received, or were threatened with, a similar treatment. About 1412 Hythe suffered from a series of disasters and the "services" of the town were remitted for the next five occasions on which the Ports might be summoned.

The Court of Shepway, or Shipway, was the chief law-court of the corporation. This "portmote," or parliament, made by-laws for the Cinque Ports as a whole and it acted as a Court of Appeal. It was held originally, and for some centuries, at Shipway Cross which is north of West Hythe and less than a mile from Stutfall Castle. Before 1597 the installations of all the Lord Wardens took place here. The meetings are believed to have been held in the open, like all early Teutonic assemblies. Ordinary business was conducted by the courts of brotherhood ("brod-hall") and guestling which generally sat at Dymchurch or Romney. The fact that Shipway Cross should have been chosen as the meeting-place of the most important gatherings is clear evidence of the importance of the Hythe district in early times.

According to local tradition, Hythe was ravaged by the French seven times*. There appears to be only one historical record of these incursions, however. Hythe was an unwalled town and its churches were probably the only buildings which could offer substantial protection†, but, as far as is known, it suffered far less from pillage than did Winchelsea and Rye which were well defended. In 1295 a French fleet of 300 ships, drawn principally from Mediterranean ports, sailed up the Channel. One vessel got ahead of the others while exploring the land and it grounded near Hythe. After enticing the crew a short way inland, the townsmen turned on them and killed them to a man. There were 240 foreigners slain and their ship was burnt‡. The remainder of the fleet withdrew, but it returned later and burnt Dover before being finally repulsed. The earliest account of the bones in St Leonard's Church mentions a tradition that they are those of Frenchmen slain on the coast, and it is

but it was written for the general reader and there is not a single exact reference in the whole volume. It is to be regretted that Professor Burrows did not incorporate the results of his researches in a more technical form. Ford Madox Hueffer, *The Cinque Ports*, 1900. This is described by the writer in the preface as "a piece of literature pure and simple" and, though presenting some new facts, it is far from accurate in many details. There seems to be a real need for a comprehensive history of one of the most important English institutions of mediaeval times.

* This statement is made by Hueffer (*op. cit.*, p. 193). The writer makes another statement for which we have not been able to find any confirmatory evidence. He says (p. 192) that for the whole duration of the cult of St Thomas, Hythe was a principal port of entry for foreign pilgrims. St Thomas was believed to have declared oracularly that Hythe was the safest port for those sailing to Boulogne. The town was ecclesiastical property at this time.

† In the town records for 1412 mention is made of carrying guns from the bridge to the Church on sleds (*Fourth Report of the Royal Commission on Historical Manuscripts*, 1874, p. 434).

‡ The story is given by Henry Knighton (*De Eventibus Angliae*) who wrote in the following century.

probable that this occasion is referred to. A later account refers to it directly and suggests that the collection was started in this way, though there must have been additions to it later (see pp. 155—157 below).

The primary cause of the decline of Hythe was undoubtedly the silting-up of its harbour. There is a considerable amount of evidence bearing on this matter and it has been interpreted in different ways. The theories relating to the nature of the coast in Roman times have been considered above and the most plausible one postulates the existence of a pool harbour extending from a point below West Hythe to Seabrook. There were centres of population at this time at West Hythe and where the present town of Hythe stands, which would have been the middle of the harbour. From the fourteenth century onwards many attempts were made to enlarge and prevent the further choking of this pool*. Leland visited the town sometime between 1535 and 1543, and he gives the following account of it:

"Hithe hath bene a very great towne yn lenght, and conteyned iiii paroches that now be clene destroied, that is to say S. Nicolas paroche, our Lady paroche, S. Michael's paroche, and our Lady of Westhithe, the which is with yn lesse than half a myle of Lymme Hille. And yt may be well supposed that after the haven of Lymme, and the great old town ther fayled, that Hithe strayte therby encresed and was yn price. Finally to cownt fro Westhyve to the place wher the substans of the towne ys now ys ii good myles yn lenght, al along on the shore to the which the se cam ful sumtyme, but now by bankinge of woose and great casting up of shyngel the se ys sumtyme a quarter, sumtyme *dim.* a myle fro the old shore. In the tyme of King Edward the 2, there was burned by casuelte xviii score howses and mo, and strayt folowed great pestilens, and thes ii thinges minished the town. There remayne yet the ruines of the chyrches and chyrch yardes. It evidently apereth that wher the paroch church is now was sumtyme a fayr abbay. Yn the quire be fayre and many pylers of marble, and under the quier a very fair vaute, also a faire old dore of stone, by the which the religius folkes cam yn at mydnight. In the top of the chirch yard is a fayr spring, and therby ruines of howses of office of the abbey; and not far of was an hospital of a gentilman infected with lepre....The havyn is a prety rode, and liith meatly strayt for passage owt of Boleyn (Boulogne). Yt croketh yn so by the shore a long, and is so bakked fro the mayn se with casting of shinggil, that smaul shippes may cum up a larg myle toward Folkestan as yn a sure gut†."

"From Hithe to Holde Hithe, alias West Hithe, about 2 myles, Mastar Twyne saythe that this was the towne that was burnid alonge on the shore, where the ruines of the churche yet remayne‡."

* References to this matter are given in the *Fourth Report of the Royal Commission on Historical Manuscripts*, 1874, pp. 429—439. A selection of the town records is printed in this volume; the majority of the others have not been published. They relate principally to the fifteenth century and are by far the most important collection illustrating the history of Hythe as a Cinque Port and its municipal and general condition at that time.

† *The Itinerary of John Leland in or about the Years 1535—1543.* Edited by Lucy Toulmin Smith, 1909, pp. 64—65.

‡ *Ibid.*, p. 46.

Writing in 1570, Lambarde declares that he could find nothing to account for the early greatness of Hythe. He concludes that: "either the place was at the first of little price and for the increase thereof endowed with privileges or (if it had been at any time estimable) that it continued not long in that plight*." The earliest port is said to have been at Lympne, but the sea deserting that part it moved to West Hythe until it, too, became land-locked and Hythe took its place: "which now standeth indeed, but yet without any great benefit of the sea, forasmuch as at this day the water floweth not to the town by half a mile or more."

Writing shortly after Lambarde, Camden in 1586 gives a similar account of the matter. "Nor," he writes, "is it very long since its (Hythe's) first rise, dating it from the decay of West Hythe, which is a little town hard by to the west, and was a haven, till, in the memory of our grandfathers, the sea drew off from it; but both Hythe and West Hythe owe their origin to Lime, a little village adjoining, formerly a most famous port before it was shut up with the sands that were cast in by the sea†."

The date when the harbour ceased to be used can be determined with some certainty. In the Jurats Account Book for 1412 there are frequent references to a "Hollandyr" who was evidently employed in order that he might survey and preserve the harbour. According to a survey made in 1566, there were: "creeks and landing-places 2; th'on called the Haven, within the liberties; th'other called the Stade, without the liberties. It had of shipping, 17 tramellers of five tunne, seven shoters of 15, three crayers of 30, four crayers of 40; persons belonging to these crayers and other boats, for the most part occupied in fishing, 160‡." The creek must have become too small to be navigated by any but the smallest vessels shortly after this date as the contemporary maps suggest. Some attempts to keep it open were still being made more than a century later, however. Dr Wallis wrote in 1701: "At Hythe in Kent (which is one of the Cinque Ports) there was (in our fathers time) a convenient harbour for small vessels; which is now swarved up. Several attempts have been made to recover the harbour, but with small success§."

The question when the port moved from Lympne to West Hythe and from West Hythe to Hythe, if this transference ever did take place, is obviously one which cannot be answered so decisively. Camden says that the second of these movements took place not "very long since." Having, apparently, no better evidence to go on than the three sixteenth century accounts quoted above, Hasted concluded that West Hythe was once part of Hythe itself and that the two were connected by a number of straggling houses. He also adopts another suggestion made by some

* William Lambarde: *A Perambulation of Kent: conteining the Description, Hystorie, and Customes of that Shyre, collected and written (for the most part) in the yeare 1570*, 1576, p. 159.

† William Camden: *Britannica Antiqua.*

‡ This passage is contained in a survey of the maritime parts of Kent made in the eighth year of Elizabeth's reign. It is quoted by Hasted (*op. cit.*, Vol. III. pp. 412—413) from the Dering MSS.

§ "Of the Isthmus between Dover and Calais," *Phil. Trans.* No. 275. The "stade" was used for beaching boats after 1700, as at Hastings, but there was nothing which could possibly be called a harbour by then.

earlier writers and states as a fact that West Hythe was the original Cinque Port. These ideas have been adopted, or elaborated, by many of Hasted's successors: some of them suppose that West Hythe was the original Port and most write confidently of a town two miles in length*. These conjectures appear to be quite fanciful. There were once at least five churches in the locality and all these were probably built in early Norman times. Three, including St Leonard's which was almost certainly the largest, were situated within the boundaries of modern Hythe, one was mid-way between Hythe and West Hythe and the last, which was certainly small, was at West Hythe. Leland writes, "Holde Hithe, alias West Hithe," and "East Hithe" was occasionally used by other writers, but the general custom from earliest times was to refer to these places by the names which they bear to-day. Lewin says: "I cannot find any authentic record that West Hythe was ever anything more than a suburb of Hythe†." This appears to us to be the only safe conclusion to draw from the evidence, though we should not use the word suburb as it is extremely unlikely that the two were ever connected by buildings. It is practically certain that West Hythe was not the original Cinque Port. With its favourable position in the middle of the harbour, the present site of the town, where there is known to have been a Roman settlement, may well have become the most important centre in the neighbourhood from the time at the end of the occupation when Stutfall Castle was deserted.

The lords of Hythe in mediaeval times were the monks of Christ Church and the Archbishops of Canterbury. Professor Burrows remarks that this town and Romney were the nautical outlets "for a body of settlers scattered over a considerable space... who were for the most part tenants of the monks endowed by the Kings of Kent, rather than the centres of a fishing and trading people gathered round the ports which gave access to the interior of the island and linked it with the Continent‡." One would expect to find in such a town that the number of ecclesiastical foundations was out of proportion to the size of the population. Leland says that a "fayr abbay" once stood where the parish church (St Leonard's) is now, but nothing whatever is known about this and it is extremely probable that it never existed. The ruins in the churchyard may have been those of other buildings, but there is no other record referring to them. The sites of the churches of the "iiii paroches that now be clene destroied" have all been identified and there may have been another one§. Two—or three, if the doubtful one is included—are practically within the modern limits of the town and hence close to St Leonard's. Another is midway between Hythe and West Hythe and the last is at West Hythe. This last church is in ruins to-day and it is the only one of which there are any remains still above ground. It had fallen into disuse before Leland's time and all the others were in ruins by 1400. The church at West

* That the town was once of such a size is the generally accepted view to-day. It was adopted by Professor Burrows (*op. cit.*, p. 84).

† *Loc. cit.*, 1866, p. 366.

‡ *Op. cit.*, pp. 44—45.

§ Rev. G. M. Livett: "West Hythe Church and the Sites of the Churches formerly existing at Hythe," *Archaeologia Cantiana*, Vol. xxx. 1914, pp. 251—262. Several references to the demolished churches will be found in the records printed in the *Fourth Report of the Royal Commission on Historical Manuscripts*. The sites of two of these in Hythe are shown on the 6-inch Ordnance map.

Hythe is known to have been small and it is probable that St Leonard's was always the largest and most important in the district. There is said to have been no burial-ground at West Hythe; one of the disused churches in Hythe (St Nicholas) was surrounded by a graveyard from which bones have been taken in recent years and there seems to be no sufficient evidence to show whether the others had grounds of their own, or not. It is not known when the leper Hospital of St John was founded, but it was in existence before 1336. In 1562 it contained eight beds. In 1336 Hamo, Bishop of Rochester and a native of Hythe, founded the Hospital of St Andrew for ten poor persons. This was transferred to another site in 1342 and it became known as the Hospital of St Bartholomew. Thirteen poor persons were housed in it then and there is believed to have been a chapel, and possibly a burial-ground attached. The two hospitals are used to-day as almshouses.

The Church of St Leonard was founded in early Norman times and it was enlarged considerably in late Norman times and again in the thirteenth century. It is one of the finest parish churches in England with an inside length, excluding the tower, of 120 ft.; it has north and south transepts. The chancel is particularly large and its floor is raised to an unusual extent above that of the nave. It was built in the early thirteenth century and this plan is believed to have been adopted in order that a vaulted passage might be formed under the east end of the church to provide an ambulatory or processional way outside it. The new chancel extended right up to the road on its east front and the only possible method of constructing a path on consecrated ground, allowing passage from the south to the north side outside the building, was the one adopted. In his description of St Leonard's Church*, Canon Scott Robertson gives other examples of churches having exterior processional ways which modified the construction of the building. He also cites various orders for processionals issued by the Crown in the fifteenth century and other records relating to the matter. The ambulatory of St Leonard's has no connection with the enclosed space below the rest of the chancel and there are no windows in it. The church is built on the side of a hill rising to the north and there is a built-up path along its south exterior bounded by a wall which falls to the road below. The floor of the ambulatory is several feet above the level of this road, but it is below the level of the ground on the north side of the church. Earth had accumulated there right up to the apex of the doorway until it was cleared away in recent times. Being really above the level of the ground, this chamber, which became the ossuary, cannot properly be called a crypt. Dealing with the question of when the use for which it was built no longer existed, Canon Scott Robertson concludes: "As the Procession way would be in constant use until the Reformation was fully established, I feel confident that nothing would have been allowed to obstruct free passage through it with cross erect in solemn procession before the Reformation. Consequently I believe that the large collection of human skulls and bones now stored

* "St Leonard's Church, Hythe," *Archaeologia Cantiana*, Vol. XVIII. 1889, pp. 403—420. An interesting account of the church and town is given by the Rev. Herbert D. Dale, sometime Vicar, in: *St Leonard's Church, Hythe, from its Foundation with some Account of the Life and Customs of the Town of Hythe from ancient Sources*, 1931.

there could not have been placed within this Procession Path until after the Reformation in the sixteenth century*." Professor Parsons suggests that the ambulatory was used as such after it had become customary to place bones in it†. The maximum width of the vault is eleven feet and it is clear that a pile as large as the one that is there to-day would have been a real obstruction. No great width appears to have been needed for the Procession, however, as the south porch of the church was built about a hundred years later than the chancel, and the east and west doors in it which originally gave access to the Path were made quite small. Professor Parsons also points out that pottery and other relics of pre-Reformation date were found near the bottom of the pile when it was re-stacked in 1908. This evidence is of little value, however, since mediaeval pottery had been found when the bones were re-stacked on earlier occasions and there is no record of whether the potsherds were in or beneath the pile. The real objection to the theory that no human remains were placed in the vault until after the Reformation is the fact that this would allow no more than a century for the accumulation of parts of over 4000 skeletons, this number having been ascertained in 1908 by counting the femora present. If the use of the vault for ritual purposes is supposed to have been discontinued about 1550, then in little more than 100 years afterwards (see p. 155 below) the stack of bones must have been nearly completed and it became the regular custom to show them to visitors. The yearly deaths in the town for this century are known to have been between 30 and 40. It is unlikely that the grave-diggers would disturb and preserve part of one skeleton every time they dug a grave and it was certainly only on rare occasions that they placed bones of children in the vault, although they must have dug them up quite frequently. On Canon Scott Robertson's hypothesis, it is only possible to reconcile these numbers by supposing that at some particular date between 1550 and 1650 there were at least 1000 skeletons exhumed at one time. These may have come from the graveyard of St Leonard's, or from that of one of the demolished churches, but there is no record of such an event and we have no reason to believe that any of the disused graveyards were built on during this period.

The original area of the burial-ground of St Leonard's Church was probably quite small. After the chancel was built, it must have extended principally from the north side of the church, partly from the west also, but not at all from the south or east fronts as there were roadways along these two sides. The oldest tombstone found to-day in the area mentioned was erected in 1649, but the majority there date between 1780 and 1850. After the last of these dates the area was apparently extended to the north and west, and about 1880 a piece of land was added to the east of the church. Interments now take place in a ground removed from it. The area of the graveyard before 1700 was probably less than one acre. The Early Norman church was extended 27 ft. to the east when it was enlarged in Late Norman times and there was a further extension in the same direction, and to the same extent, when the present chancel and ambulatory were built in the early

* *Loc. cit.*, pp. 406—408.

† "An explanation of the Hythe Bones," *Archaeologia Cantiana*, Vol. xxx. 1914, pp. 203—213.

thirteenth century. It is extremely probable that the ground built on at these times formed part of the original graveyard, and not at all unlikely that the bones were disinterred then and that they were placed in the ambulatory as soon as it had been constructed. This was about 150 years after the foundation of the church, and it cannot be supposed, of course, that any number approaching 4000 skeletons can have been found then in a part only of one graveyard while other grounds were in use in the town. It appears to us most probable that a nucleus of the collection was formed in this way and that additions were made to it by grave-diggers during the following centuries. According to this hypothesis the passage must have been used at the same time both as an ambulatory and as an ossuary. The alternative one which supposes that no bones were placed there until the Reformation cannot be disproved and the possibility that part, at least, of the remains were taken from the other graveyards or from plague-pits must be admitted. The evidence is not sufficient to decide such questions definitely*.

The church registers have not been published. The word "Plague" is written against the entries for the years 1597 and 1623. The following totals refer to complete decades:

Years	Burials	Marriages
1587—1596	298	84
1597—1606	309	119
1613—1622	433	?†
1623—1632	434	106
1633—1642	361	61

During the Commonwealth there were hardly any entries in the registers. The survey made in 1566 referred to above gives the number of inhabited houses as 122, the number of persons lacking habitation as 10 and the total engaged in fishing 160. There are two maps of Hythe and its immediate surroundings drawn in 1684 and 1685 respectively by Thomas Hill, "sworne Surveyor in Canterbury." The first was prepared with the object of showing the land belonging to St John's Hospital and the part showing the town is reproduced in Fig. 2‡. There appear to be about 260 separate buildings and the area covered is approximately the same as that occupied by Hythe to-day if the buildings of the School of Musketry and all south of the Royal Military Canal are left out of account. The 1685 map shows the town on a small scale and there are far fewer buildings: it shows the lands belonging to the two hospitals of St John and St Bartholomew. We are indebted to the Governors of these institutions for permission to examine both maps and to reproduce part of the earlier one. The surveyor was not concerned principally with the town, and the number of houses shown on the 1684 map may be inexact as some other particulars certainly are. The census of 1801 gives a total of 212

* [If a charnel-house existed in the graveyard as in so many English graveyards before the Reformation, its contents, including possibly the bones disinterred when the ambulatory passage was built, may have been transferred there after the Reformation, because either the charnel-house was dilapidated, or its site was required for other purposes. Ed.]

† There were 76 marriages in the eight years 1615—1622.

‡ A reproduction of the complete map is given in *Archaeologia Cantiana*, Vol. xxx. 1914.

Fig. 2. A map of Hythe drawn in 1684 by Thomas Hill and reproduced by permission of the Governors of the Hospitals of St John and St Bartholomew, Hythe. The site of the church and churchyard of St Nicholas is shown to the west of the town near the spot marked H.

inhabited houses in the parish of St Leonard and a population of 1365. In 1800 there were 43 burials, and at the beginning of the eighteenth century there were generally fewer than 20 burials a year. The evidence suggests that between Elizabethan times and 1800 there were certain fluctuations in the size of the town, but it was probably larger at the later date than at any other time during the period. It is extremely probable that the population was rather larger than this in the earlier centuries, but it is not clear that it was many times larger as some writers have implied. The evidence of the Domesday survey has been considered. According to Leland there were 360 houses burnt down in the reign of Edward II. No confirmation of this statement can be found and it is extremely unlikely that the catastrophe was at West Hythe as Master Twyne stated. Events which had a profound effect on the town, and which are well documented, took place about a century later. Writing in 1570, Lambarde gives an account of a conflagration which destroyed 200 houses, a pestilence and a calamity at sea which resulted in the loss of five ships and 100 men, all occurring about the same time in the reign of Henry IV *. The town was so badly hit that the survivors decided to abandon it, which they would have done if the King had not granted them a charter releasing them from services on the next five occasions on which the Cinque Ports should be summoned. Lambarde says that he had seen this charter. It is among the archives of the Corporation of Hythe referred to by H. T. Riley † and it is dated in the second year of the reign of Henry V (1414). The catastrophes are proved by other documents to have taken place in the time of Henry IV and all other particulars related by Lambarde appear to have been correctly reported. Records of the early fifteenth century represent the town as being in a state of utter filth and disrepair and, owing to its declining harbour, it appears to have had insufficient vitality to recover the important position it must have held in the thirteenth and fourteenth centuries.

We should certainly not expect to find that there were many foreigners in the town at the time when there was a long-continued internecine feud between the Cinque Ports and the French fleets. Shortly after this, however, a Hundred paper (for 1422) records a Jurors' report that there were many Frenchmen in the town who had not taken the oath to the King ‡. The Patent and Close Rolls of earlier years refer to orders which were given at different times to encourage, or obstruct, foreign traders. It is known that considerable numbers of foreigners settled in Kent in the reign of Edward III (1327—1377), but these were of Flemish or Walloon origin and none of them would have been called Frenchmen at this time. The influx of Huguenot refugees did not commence until nearly two hundred years later, and large numbers of them are known to have settled in the Kentish ports, and in Winchelsea and Rye in particular. Hythe does not seem to have received any of them, however. According to a census taken in 1622 there were only three

* *Op. cit.*, pp. 141—143.

† *Fourth Report of the Royal Commission on Historical Manuscripts*, 1874, pp. 429—439.

‡ This is quoted on p. 432 of the Report of the Royal Commission cited above.

foreigners in the town and these came from Flanders*. Few foreign names occur in the town records, or, it is said, in the parish registers, and there is no evidence that the nature of the population of Hythe was modified at all by admixture with aliens any time after the Norman Conquest.

We may now consider the earliest direct references to the bones and the various hypotheses which have been advanced to account for their origin. Leland (1535—1543)† gives a description of the town and he mentions the "very fair vaute" under the choir of the parish church, but nothing is said about its contents. Lambarde (1570), Camden (1586) and Kilburne (1659) do not mention the ossuary. It is particularly unfortunate that Hythe should have been omitted entirely from Thomas Philipott's *Villare Cantianum* (1659), and it is not known why this striking omission occurs. The work is supposed to have been compiled principally from materials collected by John Philipott—the father of the author and a native of Folkestone—before 1640. The discovery of the crypt below Folkestone Church is referred to (see p. 142, and second footnote there). What appears to be the earliest description of the skeletons is one given by Jeake in his annotations to the Charters of the Cinque Ports‡. These had been written by 1678 although not published until 1728. The following is the complete account:

"On the north (*sic*) side of the church is a charnel-house, or *Golgotha*, full of dead men's bones, piled up together orderly, so great a quantity as I never saw elsewhere in one place; supposed by some to be gathered at the shore, after a great sea fight and slaughter of the French and English on that coast, whose carcases, or their bones after consumption of the flesh, might be cast up there, and so gathered and reserved for a memorandum."

Someone before Jeake had associated the skeletons with a skirmish between the inhabitants of the town and a party of Frenchmen, and the bones must have been seen by several people some time before he wrote. There is no suggestion that any additions were being made to the orderly pile of bones at this date. Jeake was Town Clerk of Rye and he was certainly well acquainted with local affairs. James Brome, our next authority, was Rector of Cheriton, near Folkestone, from 1676 to 1719, and he was also Chaplain to the Cinque Ports and Vicar of Newington, which is 2½ miles north-east of Hythe. His book was published in 1700, but the materials for it had been collected some years earlier. After a short description of the town, Brome writes:

"But which now more especially preserves still the fame, and keeps up the repute of this poor languishing port...is the charnel-house adjoining to the church, or the arched vault under it, wherein are orderly piled up a great stack of dead men's bones and skulls, which appear very white and solid, but how or by what

* *Lists of Foreign Protestants and Aliens resident in England, 1618—1688, from Returns in the State Paper Office*, Camden Society, 1862, p. 14.

† The dates given here are the latest years, or periods, in which the several writers can have visited Hythe before writing their accounts of it, as accurately as these dates can be determined. References have been given above unless otherwise stated.

‡ *Op. cit.*, p. 109.

means they were brought to this place the townsmen are altogether ignorant, and can give no account of the matter...." The writer supposes that they are probably those of the Frenchmen killed in 1295 and "after this slaughter these men's bones in all probability might be gathered up and laid there after which daily accessions of more might be made till they encreased to so vast a number as is still visible*."

At some time before 1700, and probably from about 1650, it had evidently become the custom to show the bones to visitors and this has been continued without intermission until the present day. No later account of the town has omitted to comment on its most spectacular attraction. It has been suggested that the earlier topographers fail to mention the bones because similar collections could then be seen at many other churches. It is known that there were ossuaries in the parish churches of Folkestone, Dover and Upchurch in Kent, and as late as 1751 corpses were being deposited in crypts in the same county†. It is not known, however, that it was customary to show these collections to visitors, and it is clear that this cannot have been the common practice at all the churches which possessed ossuaries. The evidence suggests forcibly that a large pile of bones had been gathered together in the ambulatory of St Leonard's by the middle of the seventeenth century and that about this time they were shown regularly. It is unlikely that there were many later additions to the pile. There is an obvious reason why the bones here should be shown while the ossuaries of other churches were being bricked up and forgotten. Unlike them it is entirely above ground and, owing to its position, it is adequately lighted and ventilated.

The next writer of any importance who discusses the origin of the skeletons is Dr John Harris whose *History of Kent* was published posthumously in 1719. He says that it had been a "long and very common enquiry, how and on what occasion they came there." Three conjectures are offered: the first is that they are the bones of people buried in the grounds of the four ruined churches, and the fact that similar collections are found in other churches lends support to this view; the second is that they are the remains of the Frenchmen slaughtered in the town in 1295, and the third that they are the remains of the Saxons and Britons who fell in the last battle which Vortimer had with the Saxons. The writer is inclined to think that the last is the true explanation, and the fact that another ossuary had been discovered at Folkestone Church is supposed to render it more probable. The account given by Charles Seymour in 1776 need only be mentioned because it appears to contain the first printed reference to an inscription hanging in the vault attributing the remains to the Danes who are supposed to have been slain in the neighbourhood. It is not known who originated

* *Three years travels over England, Scotland and Wales*, p. 270.

† An Act of Parliament (25 George II c. 11) was passed in that year: "To enable the parishioners of East Greenwich to deposit Corpses in the Vaults or Arches under the Church of the said Parish and to ascertain the fees that they shall pay for the same."

Other ossuaries exist, or at one time existed, at Rothwell, Waltham Abbey and Manchester and Ripon Cathedrals, and there were doubtless many others. The coffin containing the remains of John Hunter was once housed with many others in the crypt of the church of St Martin-in-the-Fields and bones not in coffins have been found there.

this theory, or when the inscription was first exhibited. Hasted (1790) rejects it and adopts a variation of one of Harris's suggestions. The battle in 456 was fought somewhere between Hythe and Folkestone and, after the bones of the slain had remained for a time exposed on the sea shore until they became white, those of the Britons were removed to Hythe and those of the Saxons to Folkestone. The possibility that the bones might have been derived from graveyards is not even considered. One or other of the above theories has been accepted by all later writers on the subject.

Anthropologists are accustomed to finding that historical evidence has been perverted in order to establish a connection between a collection of human skeletons and a battle. In the present case there are two battles involved. They were both fought some centuries before the collection of bones could have been made and it is extremely unlikely that the field of either was anywhere close to Hythe. The authority of Hasted and the spurious evidence of the inscription, first hung up in the ambulatory about the middle of the eighteenth century, and of which an old copy may still be found there, have been responsible for perpetuating obvious errors. The suggestion that the Frenchmen known to have been slain in the locality in 1295 are represented is a far more plausible one, but there are said to have been fewer than 300 of these, and such numbers are generally exaggerated, while the imperfect remains of over 4000 skeletons have been found in the ambulatory. Brome appreciated this difficulty and he supposed that the original pile was increased by the later addition of other bones which came, presumably, from graveyards. The real objections to this view are, firstly, that such an assemblage would have shown a greater disproportion between the sexes than the one actually found and, secondly, that it would have been racially less homogeneous than it actually is. By far the most plausible hypothesis in the light of all the evidence appears to be that all the bones were taken from graveyards and it is probable that the majority of them, at least, came from the ground of St Leonard's Church. The date when the collection was started cannot be ascertained with certainty, but it may have been when the ambulatory was built in the early thirteenth century. There were probably few, if any, additions to it after the time about the middle of the seventeenth century when the ambulatory passage was first shown to visitors. There is little historical evidence bearing on the ethnic history of the inhabitants of Hythe. The district must have been almost as thickly populated as any in Kent in Roman times by marines, auxiliaries and traders. It is not known that there was any large Jutish settlement in the neighbourhood, but immediately before the Norman Conquest the town was again a relatively important one. The origin of its mediaeval population may be disclosed by a study of the physical characters of the people themselves.

(3) *Anthropological Descriptions of the Hythe Skeletons.* The earliest account of the Hythe skeletons provided by anyone capable of examining them from an anthropological point of view appears to be a short one published in 1834* by

* *Physiognomy founded on Physiology, and applied to various Countries, Professions and Individuals: with an Appendix on the Bones at Hythe....The Sculls of the ancient Inhabitants of Britain and its Invaders.* The appendix is pp. 280—286.

Alexander Walker, an Edinburgh anatomist and physiologist of repute in his day. He remarks that no rational account of the remains had been given by an anthropologist before his time. The contention that the skeletons deposited in the ambulatory were those of the soldiers slain in 456 in the battle between the Britons under Vortimer and the Saxons is accepted and the alternative one associating them with the Danish invaders of the tenth century is discarded as being less probable. The skulls are not those of one race, however, as had been supposed. "Two forms of scull, very distinct from each other, predominated:—one a long narrow scull, greatly resembling the Celtic of the present day;—the other, a short broad scull greatly resembling the Gothic....These were mixed with others of less definite character, in general so varied as to fall under no such obvious classification." The mean maximum lengths and breadths of unstated numbers of the Celtic (Ancient British) and Gothic (Saxon or Danish) types are given and there are drawings in *norma verticalis* of contrasted specimens. The report that red hair had been found led to a search, since this is "a striking characteristic of the Gothic nations." Walker writes: "I had not proceeded far, when I found several *sculls with masses of red hair matted upon them*....In every instance, *these were the short and broad Gothic sculls*; and nothing of the kind could be discerned on the British!" In addition to the two principal types of cranium, another kind was found which was round-headed, but stronger, heavier and more capacious than either of the preceding kinds. These larger skulls, however, are "universally in a more imperfect state" than the others and they must be ascribed to Romans who had been killed on the same battle-field in a previous age. The evidence of their superior development "made one cease to be surprised that the Roman was easily the conqueror and the master of the other races around him."

The next account of the Hythe skeletons worthy of notice is that of another Edinburgh anatomist, Robert Knox, who was a better known anthropologist than his predecessor in this field. He contributed two papers to the *Transactions of the Ethnological Society* in the early sixties*. The first gives a general description of the bones and of sundry anomalous specimens found. A few skulls of children were the only ones not attributed to adult men and it is stated explicitly that no female crania were observed. The type is said to resemble that of the present inhabitants of South Britain with "few varieties, and none peculiar or different from what we now find." It was concluded that anomalous conditions, such as inter-parietal bones and fronto-temporal articulation, occurred less frequently than in modern times. The various theories that had been proposed at different times to account for the contents of the ossuary are discussed, and Brome's attributing them to the French invaders massacred in 1295 is accepted. Hence they belonged "to a mixed people composed of several races, amalgamated for the time into a nation, and

* "Some Observations on a Collection of Human Crania and other Human Bones at present preserved in the Crypt of a Church at Hythe in Kent," *Transactions of the Ethnological Society of London*, New Series, Vol. I. 1861, pp. 238—245. "Some additional Observations on a Collection of Human Crania and other Human Bones, at present preserved in a Crypt of a Church at Hythe, in Kent," *Ibid.*, Vol. II. 1863, pp. 136—140.

strictly analogous to the inhabitants of Kent at that period." The suggestion that the collection is merely the remains of churchyard bones collected promiscuously is said to be untenable, though no arguments against it are brought forward. Knox says that the distinction made by Walker between round and long heads is not perceptible. He speaks of fragments of "Roman Saxon" pottery and of coarse mediaeval earthenware found a short time before in re-stacking a portion of the pile of bones and then in his possession. The second communication of this writer followed a re-examination of the remains. The churchyard theory is said to be untenable, owing to the absence of female specimens, and also because the condition of the bones is unlike that of the skeletons of interred bodies, no signs of decay being observed. We are told that: "the pile as it now stands, and the crania on the shelves were thus arranged some seventeen years ago by the son of General Frieze, who took a fancy to trouble himself with this labour. Prior to this, the entire mass lay heaped up in utter confusion upon the floor of the crypt...." A few useless measurements are given and it is remarked that the great deficiency of the type is in its frontal breadth. The view that the bones are most probably those of men who fell in battle is still regarded as the most probable, though the date of this battle has yet to be determined. The theory that the remains of Frenchmen killed at the end of the thirteenth century had been preserved is thus given up by this writer.

At some date in, or before, 1865 Frank (Francis T.) Buckland paid a visit to Hythe and he has a short article on the bones in a book published in that year*. He does not claim to have made any special inquiries regarding the history or anthropological nature of the material. He apparently saw no reason to question the statement ascribing them to the slain in the battle in Ethelwulf's reign. This was presumed to have taken place on flat ground between the old town and the sea, and we are told that: "The House of the present mayor of Hythe is built upon part of this battle-field, and in digging the foundation of it many bones were discovered, whence a name is now given to this house, more expressive than classical, viz. 'Marrow-Bone Hall'." This was possibly on the site of the fourth church, or churchyard, in the town of which the exact location is not known. The red hair and the distinctions between the skulls of Ancient Britons and Danes were pointed out to visitors in Buckland's time. Regarding the arrangement of the remains he says: "Mr Tournay, builder and clerk of the church, informed me that the bones used to lie scattered in disorder till about twenty years ago, when they were arranged in their present decent order."

Barnard Davis's catalogue of his own collection of skulls†, which is now in the Museum of the Royal College of Surgeons, gives measurements of six specimens from the Hythe ossuary. One was purchased at the sale of Mr Heaviside's Anatomical Museum in 1829, and it is said to have been obtained "by a lady, under peculiar

* *Curiosities of Natural History. Fourth Series.* The article is entitled: "Ancestral Skulls."

† *Thesaurus Craniorum*, 1867. Measurements and descriptions of the Hythe skulls are given on pp. 42 and 44—47.

circumstances": we are not told how the others were procured. Their owner had never seen the ambulatory and its contents. Of the few skulls he had examined he says: "the *condition* alone of these crania, if the evidence of form were wanting and the period of the foundation of the church itself were not conclusive, will wholly exclude the four first people named (viz. the Britons, Romans, Saxons and Danes) from any participation in their ownership....That they could have been exposed to the air and other deteriorating influences, even in this closed crypt, ever since the days of the Royal Antiquary (Leland), now 330 years, and maintain their present appearance is quite impossible. They are undoubtedly of more recent origin, and present the conformation of the modern men of Kent." They are said to resemble the skulls of modern Germans. The measurements in inches and tenths are of little value. Using the ophryo-occipital length and the maximum calvarial breadth, the cephalic indices found are 77·0, 81·9, 82·4, 85·1, 87·5 and 88·1 and the mean index is 83·7 *. Two of the six specimens are supposed female, and an unnamed medical correspondent had told Barnard Davis that both sexes were represented in the original collection. The divergencies in this and other respects between the account of the Hythe skulls given here and the earlier ones of Walker and Knox are striking enough.

A short account of the skeletons was given by the Rev. T. G. Hall, Vicar of Hythe in 1889 †. The conjectures of Walker and Knox are referred to and the bones are supposed to be those of men slain in a battle. This paper is cited here only because of a reference to an examination of the material made by a Mr Prideaux. We are told that: "He devoted ten days to the careful examination of these remains, during which he submitted to accurate measurement some 600 skulls. He told the Vicar at the time that he was of opinion that a large proportion of them were of the Celtic type, the greater part of the remainder being of the Anglo-Saxon type. Two skulls he believed to be Roman in form, and two Laps or Danes." No other account of this investigation appears to have been published.

In the same year Canon Scott Robertson contributed an article on "St Leonard's Church, Hythe" to the *Archaeologia Cantiana* ‡. In a footnote (pp. 407—408) some particulars are given relating to the skulls. At the Canon's request they had been examined by Dr Randall Davis, surgeon of Hythe. Among 723 crania, only 36 had been found with a persistent frontal suture and three of these were juvenile. Only 10 skulls were found with injuries inflicted before death and all of these were on the anterior half of the cranium. An immense diversity of size and shape was found in the material. Although not able to distinguish with certainty between male and female specimens himself, Dr Davis was told that Professor Owen had picked out many females.

* Cf. values given in Table I below. Barnard Davis says that his Hythe skulls are remarkable for their magnitude, but they may have been selected on this account. The type is actually a small one.

† "On Human Remains in the Crypt of St Leonard's Church, Hythe," *Archaeologia Cantiana*, Vol. XVIII. 1889, pp. 333—336.

‡ *Ibid.*, pp. 403—420.

It was not until nearly 30 years later that Mr (afterwards Professor) F. G. Parsons undertook a somewhat lengthy examination of the Hythe skeletons*. He says that 100 measurements on them had been taken by Dr Randall Davis in 1899, but these do not appear to have been published. Professor Parsons deals first with the history of the bones, and after considering part of the historical evidence we have presented above he concludes that the bones had been dug up from this churchyard or neighbouring ones and stacked under the church "in the way which was quite usual in pre-Reformation days." Similar collections in the crypts of other Kentish churches at Folkestone, Dover and Upchurch are referred to, and a few mean measurements of short series from the last two are used for comparative purposes. The remains at Hythe are said to represent at least 4000 men, women and children, this number having been ascertained by "counting all the heads of thighbones seen in re-stacking the whole pile which the vicar has lately had done." Several masses of hair, in which shades of red predominate, were found at this time and also some pottery characteristic of the fourteenth and fifteenth centuries, wooden platters and part of an old shoe or boot. Seven absolute measurements and two indices are given individually for 326 male, 230 female and 34 immature skulls and the means are compared with those of several English and a few other series. The form of the distribution is shown graphically in the case of three of the characters. It is concluded that the Hythe skulls resemble modern Bavarian and Würtemberg series more closely than they do the seventeenth century London ones from Whitechapel and Moorfields. The need for more data relating to the populations of other parts of England is emphasised. The average length (not otherwise defined) of 76 male femora is given as 45·1 cm. and of 79 female as 41·8 cm. These means are greater than the corresponding mean maximum lengths for the Whitechapel femora. This paper of Professor Parsons is the first rational account of the material, but from an anthropometric point of view it is clearly inadequate†. His cranial measurements and remarks on anomalies and other features provided are dealt with more fully in later sections of our paper.

The same writer gave a lecture on the bones to the Kent Archaeological Society in 1914‡. No fresh evidence had been collected in the interim, but several new theories relating to the origin of the population are considered. Two kinds of cranium are said to be found, one being the characteristic long one met with in series from London plague-pits and the other being so short that nothing like it is found elsewhere in England. The types of Saxon and English Bronze Age skulls are not present. It is supposed that the short-headed element represents "continental people who settled here in a peaceful way with their women-folk, though I confess this is mere surmise....I can find no definite account of their

* "Report on the Hythe Crania," *The Journal of the Royal Anthropological Institute*, Vol. XXXVIII. 1908, pp. 419—450.

† In his "Report on the Rothwell Crania" (*Journal of the Royal Anthropological Institute*, Vol. XL. 1910, pp. 483—504) Professor Parsons gives a sagittal type contour for 80 male Hythe crania, but the measurements from which it was constructed are not provided.

‡ "An Explanation of the Hythe Bones," *Archaeologia Cantiana*, Vol. XXX. 1914, pp. 203—213.

coming: it may have been in the days of the wool staple in the reigns of Edward I and III, when so many foreigners were welcomed into England, or it may have been later...." An alternative suggestion is that the bodies are those of Wendish or Vandal invaders who are presumed to have settled in this country in large numbers in Anglo-Saxon times. The remains in the ossuary might represent the type of these foreigners which persisted in this region until mediaeval times. Such an explanation would account for the large proportion of skulls with high cephalic indices, since the Vandals are thought to be of Slavonic origin and closely allied to modern Poles, but among other objections to it are the complete absence of the type from all Jutish cemeteries examined and also the absence of any place-names in the district of Wendish origin.

The earlier anatomists and anthropologists who examined the Hythe skeletons did little to correct the gross errors of the historians who had speculated on their age and derivation, and they even confused the problem by suggesting new and fantastic theories. Professor Parsons corrected many of the solecisms of his predecessors and the measurements he took helped him to do this. These measurements are very inadequate, however, and no excuse need be made for re-examining the material from an anthropometric standpoint.

(4) *The Nature of the Hythe Series.* Unless otherwise stated, the following remarks refer to the samples of 199 Hythe skulls examined by the present writers. Traces of reddish-brown hair were found by us, and considerable quantities appear to have been noted by earlier observers. A plait several inches in length is preserved in the showcase in the ambulatory. The hair is similar in colour and texture to that found on ancient Egyptian skulls and there can be no doubt that it has been thoroughly bleached. There is no means of determining the integumentary colours prevailing in the living population. No sign of flesh was observed, but several of the skulls had fragments of brain in the brain-box. None of these was large and only a few had to be cut in order to extract them through the *foramen magnum*. Fragments of dried brain are commonly met with in ancient Egyptian skulls, but they are always more brittle, and more deprived of their organic constituents, than are the like fragments from Hythe. On being scraped with a knife the latter have a glossy dark-grey surface.

Turning to the colour of the crania, those excavated from London graveyards or plague-pits and from the Spitalfields site have been found to be very uniform in so far as the appearance and condition of the bone is concerned. Only small quantities of hair were discovered on them; those from the Farringdon Street and Spitalfields sites contained no remains of the brain; several of the specimens (apart from those of the Spitalfields series where only one skull was affected) have green copper stains on them and all are extensively discoloured, the majority being of a dark brown shade. No trace of staining due to copper salts was found on the 199 Hythe skulls*. As far as their condition is concerned, the whole series in the

* It is worthy of record that fragments of half decayed wood were found adhering to the supra-occipitals of two skulls (753 and No. x). Professor Parsons also found some "particles of woody fibre" mixed with a few hairs adhering to the occiputs of some specimens.

ambulatory appears to be remarkably homogeneous. Nearly all the bones are of a a whitish-grey colour which is very distinctly lighter than that almost invariably observed in the case of excavated London skeletons. One specimen (1001) is a mottled reddish-brown and a few others have dark-grey or light-brown stains on the calvaria, but the prevailing colour is little darker than that which a macerated bone might acquire after lying undusted for centuries. If a body be buried, with or without a coffin, the colour which the skeleton adopts in the course of time is determined by the nature of the soil and it may be no reliable guide to the age of the interment. The ambulatory passage of St Leonard's is entered by a door on the south side which is often open and the ventilation is more than adequate at all times. The sea air probably bleached the bones, as it certainly did the hair. The majority of the skulls now on the shelves are well preserved; some show signs of wear, such as partly eroded outer tables, and a few are reduced to a friable condition. Numbers have been painted on the frontal or parietal bones and the highest of these found was 1099. Professor Parsons says that there are enormous numbers of fragmentary skulls in the great pile which was re-stacked while he was at Hythe. The femora were counted then and they are said to represent at least 4000 people. The remains other than the skulls on the shelves have been arranged in a single pile which is roughly rectangular in form and one side is against the west wall of the chamber. The three sides which can be examined are made up almost entirely of the ends of femora with a few other long-bones and crania intercalated between them. The dimensions of the pile are approximately 24 by $5\frac{1}{2}$ ft. at the base and the average height, excluding the supporting bricks, is 5 ft. It is clear that the centre, which cannot now be examined except on the top, cannot contain, in addition to the "enormous numbers of fragmentary skulls," anything approaching the complete remains of the other parts of 4000 skeletons*. A stringent selection must have been made, either when the bones were first placed in the vault, or at some subsequent date. The collectors of

* It appears from Professor Parson's measurements that the average length of the Hythe femora without regard to sex is approximately 17 inches. On this assumption, the femora which form three sides of the pile occupy almost exactly one-third of its total volume. But the remaining two-thirds is said to consist largely of fragmentary skulls and femora, so it is clear that the other bones of the skeleton are only present in comparatively small numbers. There are at present about 1100 skulls on the shelves in the north and south bays of the ambulatory. A printed notice hanging there says that 600 skulls were taken from the pile and arranged on the shelves in 1851, while 500 have been added since. According to the same notice the bones were re-stacked in 1908 to allow the passage of air underneath the pile and to preserve them from decay. Nearly 8000 thighbones were then counted besides fragments. Several writers have given the dimensions of the pile at different dates: in 1776 it was said to be $28 \times 6 \times 8$ ft. high (Seymour), in 1783, $30 \times 8 \times 8$ ft. (see Plate I*a* below), in 1790, $28 \times 8 \times 8$ ft. (Hasted), and the *Hythe and Sandgate Guide* published in 1816 describes the pile as being 28 ft. in length, 8 in breadth and formerly 8 in height, though the last measurement had by this time been reduced by about 2 ft. owing to the decay of the lower bones. The dimensions to-day are considerably less than these. The 1100 skulls have been removed since 1816 and the process of decay has doubtless continued. No dried bones are likely to remain intact long if they are near the bottom of a stack 5 ft. high. The Rev. H. D. Dale says that some skulls were allowed to be taken away before he was Vicar of St Leonard's (*St Leonard's Church, Hythe*, 1931, p. 56). Six of these are now in the Museum of the Royal College of Surgeons, but the location of the others, if they are still preserved, is unknown. Plate I reproduces some engravings of the interior of the ambulatory made at different times.

this material, which has been drawn from a minimum of 4000 skeletons, must have selected the thighbones and the skulls, while discarding in general other parts of the skeleton and probably the mandible*. The differences between the conditions of individual specimens, in so far as the texture of the bone is concerned, are not great in general. The skulls which Barnard Davis acquired had probably been selected from a sample which was itself a stringent selection of the total population concerned. Even if this had not been so, it is not clear that he was at all justified in asserting that bones so well preserved could not possibly have been in the ossuary for more than 300 years. The Spitalfields series is almost certainly of mediaeval or Roman date. Most of the skulls were broken when houses had been built on the grave-pits, or at the time of excavation, but the bones are as well preserved as the Hythe bones and they show no signs of abrasion or decay. This is remarkable when we consider that the burial-ground at Spitalfields was in a low-lying locality which was frequently water-logged† and as a rule bones would decay more rapidly under these conditions than they would in most churchyards‡. The condition of the Hythe skeletons cannot be accepted as evidence of their modern date, and Barnard Davis was probably judging by the colour as well as the texture of the bone when he denied the antiquity of the skulls in his possession§. It is not surprising to find that fewer than half of the Hythe crania are intact. The defects of many of the facial skeletons may be due to the spade of the sexton, or to the fact that these specimens were once lying near the bottom of a pile of bones eight feet high. There are numerous post-mortem fractures of the calvaria as well and many of these must have been mistaken in the past—by anatomists as well as others—for wounds inflicted during life. Among 112 male skulls we found only three with healed wounds; two (646 and 898) being small areas on the frontal bone and the other (798) having probably resulted from a sword-cut on the right parietal which reached the brain. Among 87 female specimens, two (655 and 758) have small depressions—on the frontal and right parietal bone respectively—which appear to be healed injuries. The frequency of traumatic lesions is not unusually high in this sample, although, judging from the long-bones and skulls he examined,

* The preference for the skull and thighbones as representative of the physical remains of the dead is characteristic of much mediaeval symbolism. In most ossuaries and "Gebeinhäuser" these are the only bones preserved.

† Evidence that this was so at the end of the fourteenth century has been given in the Spitalfields paper (p. 210). The following account shows that the conditions were much the same in the middle of the eighteenth century: "Before 1776...every person in Spital Square in the Liberty of Norton Falgate (*sic*) was greatly inconvenienced by the springs in the liberty, insomuch that...the water...used to be three or four feet deep in the cellars; and the servants used to punt themselves along in a washing tub from the cellar stairs to the beer-barrels to draw daily beer...." (Steven Totten: *A Humble Representation*...," 1795.)

‡ In the Museum of the Royal College of Surgeons there are a number of skulls dredged from the Thames and some of these are probably more than 1000 years old. Nearly all are in a good state of preservation. It is known, too, that objects of cloth and leather may be excellently preserved in peat-bogs for several hundred years. A district which is subject to periodic inundations would almost certainly encourage decay when the articles in question had been buried close to the surface.

§ Brome, writing at the end of the seventeenth century, says that the bones are very white and solid.

Professor Parsons concluded (without giving figures) that there was evidence of the turbulent life which the inhabitants of a Cinque Port experienced during the middle ages. He also observes that many of the skulls had earth in the brain-cavity as well as in the facial and auditory apertures, and this is again contrary to our experience. We found a good deal of earth in the facial and basal apertures, but the endocranial surfaces of the brain-boxes appeared to be almost entirely free from it. The skulls were examined carefully before filling with mustard-seed to determine the capacities and nothing was found inside them except a certain amount of dust which might have accumulated in the ambulatory and the fragments of brain already referred to. As Professor Parsons does not mention the last, it is possible that he did not examine the contents carefully and mistook fragments of brain for lumps of earth.

The sample of 199 crania described in the present paper was selected from among some 500 specimens kept on shelves in the far (north) bay of the ambulatory. Those chosen were all adult and intact, or nearly intact. The numbers in Professor Parsons' report range from 1 to 590 and all these skulls, with the exception of one (No. 132), appear to be kept at present on the shelves in the south bay near the door. The samples dealt with are distinct ones, except No. 132 which is included in both.

It is most charitable to assume, perhaps, that the examinations of the Hythe remains made by Walker and Knox were of a more cursory nature than they would lead one to believe, for otherwise it is difficult to understand how they failed to find female specimens. The sexing of the skulls offers no peculiar difficulty and it is obvious that a considerable proportion are those of women. Of the 199 adult crania which we measured, 112 (56·3 %) were supposed male and 87 (43·7 %) female. Professor Parsons dealt with 556 adults and he distinguished 326 (58·6 %) males and 230 (41·4 %) females. The disproportion between the sexes is doubtless due to the stringent post-mortem selection which would have favoured the preservation of the stronger male specimens. As far as this evidence can show, the original population may well have been a graveyard one. The statistical constants for the series dealt with in detail in the present paper are given in Table II. A comparison with the standard deviations and recalculated means found from the measurements given by Professor Parsons is made in Table I. We have not used his auricular height as it was taken from the centre of the meatus and it is hence not comparable with the biometric measurement taken from the porion. If the two series compared had been drawn randomly from the total collection, we should not expect any of the constants to show significant differences. Among the absolute measurements the only differences between the means which exceed 2·5 times their probable errors are in the case of the male B (Δ/p.e. $\Delta = 8{\cdot}0$), male B' (2·5) and female B' (3·6). The marked difference between the male calvarial breadth evidently leads to significant differences between the indices $100\, B/F$, $100\, B/L$ and $100\, B/H'$. The female indices are not differentiated. The single significant difference in variability is found in the case of the female H''s (3·0). The discordance between the male mean calvarial breadths is, however, so marked

TABLE I.

Comparison of the Measurements of two Samples of the Hythe Skulls.

Samples measured by	Sex	F	L	B	B′	H′
		Means				
Parsons Stoessiger and Morant	♂	177·1 ± ·24 (324) 176·3 ± ·40 (112)	178·6 ± ·24 (319) 177·9 ± ·39 (112)	143·5 ± ·21 (324) 146·7 ± ·34 (112)	99·5 ± ·17 (318) 98·6 ± ·31 (109)	133·5 ± ·21 (306) 134·1 ± ·32 (112)
Parsons Stoessiger and Morant	♀	170·8 ± ·29 (230) 170·8 ± ·44 (87)	171·1 ± ·28 (227) 171·4 ± ·43 (87)	139·8 ± ·21 (230) 140·2 ± ·39 (87)	96·0 ± ·18 (227) 94·7 ± ·31 (87)	127·7 ± ·25 (222) 127·2 ± ·34 (86)
		Standard Deviations				
Parsons Stoessiger and Morant	♂	6·45 ± ·17 6·34 ± ·29	6·48 ± ·17 6·16 ± ·28	5·56 ± ·15 5·35 ± ·24	4·52 ± ·12 4·77 ± ·22	5·48 ± ·15 4·99 ± ·22
Parsons Stoessiger and Morant	♀	6·60 ± ·21 6·15 ± ·31	6·34 ± ·20 5·95 ± ·30	4·69 ± ·15 5·44 ± ·28	3·98 ± ·13 4·30 ± ·22	5·59 ± ·18 4·68 ± ·24

Samples measured by	Sex	100 B/F	100 B/L	100 H′/L	100 B/H′
		Means			
Parsons Stoessiger and Morant	♂	81·0 ± ·15 (322) {83·2 (112)}	{80·3 (319)} 82·6 ± ·24 (112)	{74·7 (306)} 75·4 ± ·22 (112)	{107·5 (306)} 109·5 ± ·32 (112)
Parsons Stoessiger and Morant	♀	82·0 ± ·18 (230) {82·1 (87)}	{81·7 (227)} 81·9 ± ·28 (87)	{74·6 (222)} 74·3 ± ·25 (86)	{109·5 (222)} 110·3 ± ·38 (86)
		Standard Deviations			
Parsons Stoessiger and Morant	♂	3·88 ± ·10 —	— 3·69 ± ·17	— 3·51 ± ·16	— 4·94 ± ·22
Parsons Stoessiger and Morant	♀	3·96 ± ·12 —	— 3·83 ± ·20	— 3·44 ± ·18	— 5·18 ± ·27

TABLE II.

Constants of the Male and Female Hythe Series.

Character	Means		Standard Deviations		Coefficients of Variation	
	Male	Female	Male	Female	Male	Female
C	1456·3 ± 7·0 (110)	1318·0 ± 7·3 (83)	109·7 ± 5·0	98·5 ± 5·2	7·53 ± ·34	7·47 ± ·39
F	176·3 ± ·40 (112)	170·8 ± ·44 (87)	6·34 ± ·29	6·15 ± ·31	3·60 ± ·16	3·60 ± ·18
$F. V. L$	176·3 ± ·40 (112)	169·8 ± ·41 (87)	6·29 ± ·28	5·71 ± ·29	3·57 ± ·16	3·37 ± ·17
L	177·9 ± ·39 (112)	171·4 ± ·43 (87)	6·16 ± ·28	5·95 ± ·30	3·46 ± ·16	3·47 ± ·18
B	146·7 ± ·34 (112)	140·2 ± ·39 (87)	5·35 ± ·24	5·44 ± ·28	3·65 ± ·16	3·88 ± ·20
IOW	98·1 ± ·24 (110)	94·3 ± ·28 (86)	3·70 ± ·17	3·84 ± ·20	3·77 ± ·17	4·07 ± ·21
B'	98·6 ± ·31 (109)	94·7 ± ·31 (87)	4·77 ± ·22	4·30 ± ·22	4·84 ± ·22	4·54 ± ·23
B''	124·8 ± ·36 (102)	119·1 ± ·35 (84)	5·32 ± ·25	4·79 ± ·25	4·26 ± ·20	4·02 ± ·21
Biasterionic B	112·7 ± ·27 (97)	108·4 ± ·35 (85)	3·88 ± ·19	4·77 ± ·25	3·44 ± ·17	4·40 ± ·23
H'	134·1 ± ·32 (112)	127·2 ± ·34 (86)	4·99 ± ·22	4·68 ± ·24	3·72 ± ·17	3·68 ± ·19
H	134·6 ± ·31 (112)	127·9 ± ·34 (86)	4·86 ± ·22	4·74 ± ·24	3·61 ± ·16	3·71 ± ·19
OH	115·4 ± ·26 (112)	110·7 ± ·24 (86)	4·02 ± ·18	3·68 ± ·19	3·48 ± ·16	3·33 ± ·17
LB	100·5 ± ·23 (112)	95·4 ± ·25 (86)	3·58 ± ·16	3·40 ± ·17	3·56 ± ·16	3·56 ± ·18
S_1'	111·4 ± ·27 (112)	106·6 ± ·31 (87)	4·21 ± ·19	4·35 ± ·22	3·78 ± ·17	4·09 ± ·21
S_2'	108·9 ± ·37 (112)	105·1 ± ·40 (87)	5·77 ± ·26	5·53 ± ·28	5·30 ± ·24	5·27 ± ·27
S_3'	96·1 ± ·32 (112)	93·7 ± ·37 (87)	5·00 ± ·23	5·17 ± ·26	5·20 ± ·23	5·52 ± ·28
S_1	127·2 ± ·37 (112)	121·9 ± ·44 (87)	5·79 ± ·26	6·02 ± ·31	4·56 ± ·21	4·94 ± ·25
S_2	122·3 ± ·44 (112)	118·2 ± ·48 (87)	6·87 ± ·31	6·65 ± ·34	5·62 ± ·25	5·62 ± ·29
S_3	116·2 ± ·48 (112)	111·6 ± ·55 (87)	7·55 ± ·34	7·58 ± ·39	6·50 ± ·29	6·79 ± ·35
S	365·6 ± ·96 (112)	351·6 ± ·94 (87)	15·03 ± ·68	12·99 ± ·66	4·11 ± ·19	3·69 ± ·19
U	518·4 ± ·97 (112)	499·6 ± ·90 (84)	15·28 ± ·69	12·51 ± ·65	2·95 ± ·13	2·50 ± ·13
Q'	323·6 ± ·68 (112)	307·7 ± ·72 (84)	10·71 ± ·48	9·83 ± ·51	3·31 ± ·15	3·19 ± ·17
Bregmatic Q'	320·6 ± ·67 (112)	304·0 ± ·71 (85)	10·49 ± ·47	9·72 ± ·50	3·27 ± ·15	3·20 ± ·17
Broca's Q'	311·3 ± ·68 (112)	297·2 ± ·76 (84)	10·64 ± ·48	10·33 ± ·54	3·42 ± ·15	3·48 ± ·18
fml	35·6 ± ·16 (111)	35·1 ± ·16 (86)	2·44 ± ·11	2·19 ± ·11	6·86 ± ·31	6·25 ± ·32
fmb	30·2 ± ·14 (110)	29·6 ± ·14 (87)	2·25 ± ·10	1·88 ± ·10	7·45 ± ·34	6·36 ± ·33
$G'H$	69·9 ± ·27 (89)	65·1 ± ·27 (72)	3·84 ± ·19	3·42 ± ·19	5·49 ± ·28	5·25 ± ·30
GL	94·9 ± ·35 (89)	90·5 ± ·36 (70)	4·83 ± ·24	4·50 ± ·26	5·09 ± ·25	4·98 ± ·28
GB	94·8 ± ·35 (99)	90·2 ± ·34 (85)	5·17 ± ·25	4·67 ± ·24	5·45 ± ·26	5·18 ± ·27
J	134·3 ± ·35 (96)	125·3 ± ·39 (71)	5·12 ± ·25	4·85 ± ·27	3·81 ± ·19	3·87 ± ·22
NH, R	51·0 ± ·19 (111)	47·6 ± ·16 (86)	2·94 ± ·13	2·26 ± ·12	5·76 ± ·26	4·75 ± ·24
NH, L	51·1 ± ·18 (112)	47·6 ± ·17 (87)	2·85 ± ·13	2·34 ± ·12	5·58 ± ·25	4·91 ± ·25
NH'	49·2 ± ·22 (99)	46·2 ± ·19 (82)	3·24 ± ·16	2·51 ± ·13	6·59 ± ·32	5·43 ± ·29
NB	24·8 ± ·11 (108)	23·7 ± ·13 (85)	1·66 ± ·08	1·81 ± ·09	6·71 ± ·31	7·63 ± ·40
DS	12·3 ± ·12 (70)	11·6 ± ·17 (39)	1·51 ± ·09	1·61 ± ·12	12·29 ± ·71	13·91 ± 1·08
DC	21·8 ± ·17 (72)	20·6 ± ·22 (44)	2·09 ± ·12	2·21 ± ·16	9·71 ± ·55	10·70 ± ·78
DA	35·0 ± ·25 (70)	32·3 ± ·29 (40)	3·08 ± ·18	2·70 ± ·20	8·79 ± ·50	8·37 ± ·64
SS	4·5 ± ·06 (106)	3·8 ± ·07 (77)	0·98 ± ·05	1·02 ± ·06	21·84 ± 1·06	26·77 ± 1·56
SC	9·6 ± ·14 (109)	9·3 ± ·14 (79)	2·16 ± ·10	1·89 ± ·10	22·53 ± 1·08	20·28 ± 1·13
G_1	50·8 ± ·19 (96)	47·1 ± ·25 (64)	3·14 ± ·15	2·91 ± ·17	6·18 ± ·30	6·18 ± ·37
G_1'	46·4 ± ·19 (98)	43·3 ± ·22 (68)	2·72 ± ·13	2·74 ± ·16	5·86 ± ·28	6·34 ± ·37
G_2	41·7 ± ·28 (51)	39·4 ± ·22 (44)	2·94 ± ·20	2·20 ± ·16	7·05 ± ·47	5·59 ± ·40
EH	12·1 ± ·28 (47)	10·3 ± ·23 (42)	2·81 ± ·20	2·24 ± ·16	23·25 ± 1·70	21·73 ± 1·67
O_1, R	42·1 ± ·10 (111)	41·6 ± ·13 (85)	1·56 ± ·07	1·76 ± ·09	3·70 ± ·17	4·23 ± ·22
O_1, L	41·8 ± ·10 (109)	41·2 ± ·13 (84)	1·56 ± ·07	1·70 ± ·08	3·72 ± ·17	4·13 ± ·22
O_1', R	40·4 ± ·11 (82)	39·3 ± ·18 (56)	1·51 ± ·08	2·01 ± ·13	3·73 ± ·20	5·11 ± ·33
Lacrymal O_1, R	39·0 ± ·11 (73)	38·2 ± ·16 (54)	1·44 ± ·08	1·77 ± ·11	3·70 ± ·21	4·64 ± ·30
O_2, R	33·0 ± ·12 (111)	32·7 ± ·14 (84)	1·90 ± ·09	1·86 ± ·10	5·77 ± ·26	5·68 ± ·30
O_2, L	33·0 ± ·12 (110)	32·7 ± ·13 (82)	1·81 ± ·08	1·81 ± ·10	5·47 ± ·25	5·53 ± ·29

TABLE II—(*continued*).

Character	Means		Standard Deviations	
	Male	Female	Male	Female
100 B/L	82·6 ± ·24 (112)	81·9 ± ·28 (87)	3·69 ± ·17	3·83 ± ·20
100 H'/L	75·4 ± ·22 (112)	74·3 ± ·25 (86)	3·51 ± ·16	3·44 ± ·18
100 H/L	75·7 ± ·22 (112)	74·7 ± ·27 (86)	3·51 ± ·16	3·71 ± ·19
100 B/H'	109·5 ± ·32 (112)	110·3 ± ·38 (86)	4·94 ± ·22	5·18 ± ·27
100 B/H	109·2 ± ·32 (112)	109·7 ± ·37 (86)	5·00 ± ·23	5·03 ± ·26
100 $(B-H')/L$	7·1 ± ·23 (112)	7·6 ± ·26 (86)	3·68 ± ·17	3·51 ± ·18
100 $G'H/GB$	73·5 ± ·34 (80)	72·0 ± ·38 (70)	4·57 ± ·24	4·72 ± ·27
100 NB/NH, R	48·9 ± ·24 (107)	50·0 ± ·30 (85)	3·67 ± ·17	4·09 ± ·21
100 NB/NH, L	48·8 ± ·22 (108)	50·0 ± ·29 (85)	3·41 ± ·16	3·94 ± ·20
100 NB/NH'	50·7 ± ·28 (95)	51·6 ± ·35 (80)	4·07 ± ·20	4·63 ± ·25
100 DS/DC	56·8 ± ·67 (70)	57·1 ± ·96 (39)	8·31 ± ·47	8·90 ± ·68
100 SS/SC	48·0 ± ·65 (106)	41·5 ± ·77 (77)	9·92 ± ·46	10·03 ± ·55
100 G_2/G_1	82·9 ± ·51 (46)	84·3 ± ·80 (39)	6·96 ± ·49	7·38 ± ·56
100 G_2/G_1'	90·7 ± ·79 (48)	91·7 ± ·75 (42)	8·07 ± ·56	7·18 ± ·53
100 EH/G_2	28·9 ± ·64 (47)	26·2 ± ·54 (42)	6·53 ± ·45	5·21 ± ·38
100 O_2/O_1, R	78·5 ± ·32 (110)	78·7 ± ·33 (84)	4·98 ± ·23	4·50 ± ·23
100 O_2/O_1, L	79·0 ± ·31 (109)	79·5 ± ·34 (82)	4·73 ± ·22	4·53 ± ·24
100 O_2/O_1', R	81·6 ± ·40 (81)	83·2 ± ·47 (55)	5·34 ± ·28	5·13 ± ·33
100 O_2/Lacry. O_1, R	84·6 ± ·42 (73)	85·9 ± ·39 (53)	5·34 ± ·30	4·17 ± ·27
100 fmb/fml	84·9 ± ·36 (110)	84·4 ± ·40 (86)	5·56 ± ·25	5·51 ± ·28
$Oc.\ I.$	59·8 ± ·16 (112)	61·3 ± ·22 (87)	2·47 ± ·11	3·04 ± ·16
$N\angle$	64°·6 ± ·22 (89)	65°·9 ± ·28 (70)	3°·04 ± ·15	3°·44 ± ·17
$A\angle$	73°·7 ± ·18 (89)	73°·4 ± ·28 (70)	2°·51 ± ·13	3°·49 ± ·18
$B\angle$	41°·7 ± ·17 (89)	40°·7 ± ·20 (70)	2°·40 ± ·12	2°·43 ± ·14
Alveolar $P\angle$	85°·9 ± ·24 (88)	86°·1 ± ·30 (72)	3°·38 ± ·17	3°·74 ± ·21
Prosthion $P\angle$	84°·0 ± ·28 (88)	84°·2 ± ·28 (69)	3°·50 ± ·18	3°·50 ± ·20
$\theta_1\angle$	29°·4 ± ·19 (88)	27°·9 ± ·22 (70)	2°·68 ± ·14	2°·76 ± ·16
$\theta_2\angle$	12°·2 ± ·22 (88)	12°·8 ± ·27 (70)	3°·05 ± ·16	3°·32 ± ·19

that it needed inquiry. It may have arisen owing to the fact that the two series in the south and north bays are really differentiated, having as we know been shelved at different periods and hence possibly corresponding to different strata in the original stack*. Or, personal equation may account for the difference. In order to decide for, or against, the later alternative, we determined independently the calvarial lengths, breadths and heights in the case of two different samples selected at random from the series measured by Professor Parsons. The resulting means are given below, no regard having been paid to sex:

	First sample			Second sample		
	Measured by					
	Parsons (P)	Stoessiger (S)	P—S	Parsons (P)	Morant (M)	(P—M)
L	174·6 (65)	174·7 (65)	−0·1	176·4 (52)	176·1 (52)	+0·3
B	141·4 (64)	142·0 (64)	−0·6	143·4 (51)	143·6 (51)	−0·2
H'	130·3 (65)	130·6 (65)	−0·3	132·2 (50)	132·1 (50)	+0·1

* A third, but less probable, hypothesis is that in selecting the more complete skulls for measurement we have unconsciously selected the more brachycephalic which, being rounder, might on the whole be less fragile. See, however, the Note inserted at the end of this memoir.

All the differences between these means are quite insignificant and the cephalic indices (100 B/L) they lead to are all between 81·0 and 81·5, so they accord well with the values for the total sample measured by Professor Parsons. We can only conclude that our series does in fact differ significantly from his in leading to a greater mean calvarial breadth and higher cephalic index*. The variabilities of the two for these characters are not differentiated.

The male and female constants for the sample which we measured in detail are given in Table II. There are 28 indicial and angular measurements and the male and female means only differ by more than three times the probable error of the difference in the cases of 100 H'/L (Δ/p.e. $\Delta = 3{\cdot}3$), 100 NB/NH, L (3·3), 100 SS/SC (6·5), 100 EH/G_2 (3·2), *Oc. I.* (5·5), $N\angle$ (3·7), $B\angle$ (3·8) and θ_1 (5·2). A mean sexual difference of the same sign as the one now observed is generally found for 100 SS/SC, 100 EH/G_2 and *Oc. I.* and none of the other differences is markedly significant. Judging from all the measurements of shape, we may conclude that the male and female samples represent the same racial type. Comparing coefficients of variation of absolute measurements and standard deviations of indices and angles, it is found that the male variability is the greater for 33 characters, the female the greater for 39 and there is equality in this respect for 5 characters. The differences exceed three times their probable errors in the case of Biasterionic B (3·4), O_1', R (3·6) and $A\angle$ (4·5). The male and female variabilities are approximately equal as is generally found. It will be shown below that the series is as homogeneous as any other European one that has been described. The sex ratios are of the same order as those given for other English series.

Professor Parsons remarks that of 517 adult skulls which he examined 242 were above 40 at death and 275 between 20 and 40 years. No particulars relating to the ages of individual specimens are given. We attempted to obtain a rough estimate of the age constitution at death of the population by recording the state of closure of the coronal, sagittal and lambdoid sutures. Remarks on these are made in the appended tables of individual measurements and, unless otherwise stated, any one of these sutures is open for its whole length. Five other series were dealt with in precisely the same way and the results are summarised in Table III. It is known that the sutures of male skulls close normally at an earlier age than those of female skulls, and the sexual difference is well brought out by these figures. It is probable that there are also racial differences in this respect, so for present purposes we may restrict the comparison to the Hythe and the three London series. The one from Farringdon St. represents a graveyard population of the seventeenth century which may have included a number of individuals who

* This point is obviously an important one as it may influence any conclusions we may reach regarding the racial relationships of the Hythe population. It has been found in the case of several other series that the index derived from the maximum length and breadth of the horizontal type contour differs from the cephalic index found from the calliper measurements of the same skulls by a small amount which is generally less than 0·5. For our 112 male crania the contour index is 82·3 and the cephalic 82·6, and for the 87 female the first is 81·8 and the second 81·9. This confirms the fact that the male cephalic index is greater than the female, although for most series there is a sexual difference of the opposite sign.

TABLE III.

The Age Constitutions of the Hythe and other Series estimated from the State of Closure of the Principal Calvarial Sutures *.

	Series	Sutures open	Sutures beginning to close or partly closed	All sutures closed	Total examined
♂	Farringdon St. Londoners ...	19 (12·2 %)	100 (64·1 %)	37 (23·7 %)	156
	Whitechapel Londoners ...	21 (15·4 %)	85 (62·5 %)	30 (22·1 %)	136
	Spitalfields Londoners... ...	103 (19·3 %)	301 (56·3 %)	131 (24·5 %)	535
	Hythe	23 (20·9 %)	62 (56·4 %)	25 (22·7 %)	110
	26th—30th Dynasty Egyptians	74 (37·0 %)	103 (51·5 %)	23 (11·5 %)	200
	Negroes (Teita)...	12 (24·5 %)	21 (42·9 %)	16 (32·7 %)	49
♀	Farringdon St. Londoners ...	94 (46·3 %)	80 (39·4 %)	29 (14·3 %)	203
	Whitechapel Londoners ...	66 (46·8 %)	49 (34·8 %)	26 (18·4 %)	141
	Spitalfields Londoners... ...	121 (49·4 %)	94 (38·4 %)	30 (12·2 %)	245
	Hythe	51 (59·3 %)	26 (30·2 %)	9 (10·5 %)	86
	Negroes (Teita)...	38 (62·3 %)	19 (31·1 %)	4 (6·6 %)	61

died in the 1665 plague. The Whitechapel sample also represents seventeenth century Londoners who were interred in what was either a clearance pit or a pest-field. The Spitalfields skulls are of unknown date, though certainly earlier than the others, and they are probably those of victims of a plague or other catastrophe. The Hythe series almost certainly represents a graveyard population. It is surprising to find that the age constitutions of all four series are closely similar. We should expect to find that the relative sizes of the age-groups would be very different between a group of people dying from natural causes and another which was subjected to an extraordinary mortality owing to a pestilence or massacre †. The fact that such differences are not found in the case of the four English series may mean that both natural and calamitous causes accounted for the deaths of certain proportions of the individuals in each of the samples.

* References to papers describing the London series are given in the Spitalfields paper: the Egyptian skulls were selected at random from the long series dealt with by Professor Karl Pearson and Miss A. G. Davin in Vol. XVI of *Biometrika*, and the negro have been described by Miss Kitson in Vol. XXIII of the same *Journal*.

† The first is known in modern times from the Registrar-General's returns, and for 1921 there were 34·1 % of the males dying over 20 who were between 20 and 55 years of age at death and 65·9 % over 55. The corresponding female percentages are 30·7 and 69·3. A thoroughgoing pestilence or massacre which did not discriminate between the ages of the victims would give age-groups similar to those of the census population. For 1921 there were 78·4 % between ages 20 and 55 and 21·6 % over 55 for the males, and for the females the percentages are 77·6 and 22·4. There would be more than twice as many in the younger group for the calamitous deaths as for the normal graveyard population. The disproportion may not be as great as this between a pestilence and a normal sample, since plagues may be expected to discriminate to some extent between the ages of the victims, but a distinct difference would still be anticipated.

In 1912 some bones were discovered accidentally on the site of the vanished Church of St Nicholas at Hythe. Three of these were removed to the ambulatory of St Leonard's and the Rev. H. D. Dale, then Vicar, says that they were quite brown when disinterred and that they have gradually become whiter since*. We had concluded before reading this account that the peculiar whiteness of the bones in the ossuary was due to the fact that they had been exposed to sea air there. The skulls from the ground of St Nicholas are light brown, and this appears to distinguish them from all in the original collection except No. 1001 which is similarly discoloured. Only one, which is probably a female, is complete enough to give the cephalic index and this has the high value of 87·9. The type appears to be very similar to that of the long series. Among the skulls which we examined in detail there are a few exceptional ones which bear a more or less close resemblance to the type with a peculiarly low vault and markedly retreating frontal bone which is found frequently among the three seventeenth century London series. Such are the male specimens 132, 772, 808, 823 and 825 and the female 660, 806 and 849. There is no sufficient justification, however, for supposing that these skulls represent a different race from the others. It seems that there is not a single specimen in the whole series which need be segregated on that account.

Since the Hythe is an English series, its most marked peculiarity is its high cephalic index. If the present-day population of the town is descended in the main from that earlier one then we should expect to find that its mean index is also peculiar. No head measurements of Kentish people appear to have been published. We are indebted to the Headmaster of the county school at Hythe for permission to measure the head length and breadth of 49 boys between the ages of 6 and 14. These were selected as being, as far as possible, representative of the families resident in the town for several generations†. Their mean cephalic index is $79·57 \pm ·29$ and the standard deviation is $3·03 \pm ·21$. The last is not an unusual value and the sample may be supposed as homogeneous as most living ones which have been accepted as representing a single racial type. It is known that the index is uncorrelated with age and it may be compared with several means which have been given for English series from other parts of the country‡. The value found for 2313 boys was $78·87 \pm ·045$ and the mean for the Hythe boys differs from this by less than 2·5 times the probable error of the difference. The means for the longest male English living series available, comprising 9 based on 100 or more individuals, all lie between 78 and 79. The index for the living head is known to be about two points greater than that for the skull, and for the seventeenth century London series from Whitechapel, Farringdon St. and Moorfields the male values found are 74·3, 75·4 and 75·5 respectively. The living and dead populations are thus similar to one another in this respect, although the former appears to conform

* *St Leonard's Church, Hythe, from its Foundation, with some Account of the Life and Customs of the Town of Hythe from ancient Sources*, 1931, p. 62.

† Boys from the barracks are also taught in the school but none of these was measured.

‡ Karl Pearson and L. H. C. Tippett: "On Stability of the Cephalic Indices within the Race," *Biometrika*, Vol. XVI. 1924, pp. 118—138. The fact that the cephalic index is uncorrelated with age after the fifth year, at least, is demonstrated in the above paper.

to a type which is slightly more brachycephalic than that of the latter. The Hythe series has a mean cephalic index of the order 82 which would correspond with one of about 84 for the living. Instead of this we find a value of 79·6. As far as can be told from this evidence, only a small part of the present-day population of Hythe is descended from the mediaeval population of the town, and the modification which the type has undergone has been in the direction of that one to which almost all modern Englishmen conform.

(5) *Remarks on Anomalies.* The Hythe skulls were examined for anomalies in exactly the same way as the Spitalfields have been. Our samples comprised 112 male and 87 female specimens, all of which are nearly complete, and, unless otherwise stated, it may be assumed that these totals were available for examination in the case of a particular anomaly. A few anomalous skulls are shown in glass cases in the ambulatory of St Leonard's Church, but, since the series examined in detail forms only a small part of the total there, this fact is not likely to have influenced our frequencies to any appreciable extent.

(*a*) *Sutures.* Remarks on the state of closure of the coronal, sagittal and lambdoid sutures are appended to the tables of individual measurements. If no remark is made the suture is open for its whole length. The rough appreciation of the age distributions for the two sexes which can be derived from this evidence has already been considered. The sagittal suture was generally the first, or one of the first, to close. An exception to this rule was noted only in the case of one male skull (No. 1042), which has the coronal obliterated, the sagittal closed and the lambdoid closing. There are five exceptions among the females (Nos. 692, 745, 758, 832 and 899). Only one such case, and that a male showing a definitely anomalous state of synostosis, was noted for the Spitalfields series. Professor Parsons states that 6·5 °/₀ of the Hythe skulls he examined were scaphocephalic, but he remarks that he was unable to associate this condition with premature closing of the sagittal suture, and it is hence not clear how he has defined the anomaly. No example of scaphocephaly being due to the sagittal suture closing well before the others occurred in our sample. It was usual for the coronal and lambdoid sutures to begin closing after the sagittal and at approximately the same time as one another. The following specimens have complete, or nearly complete, metopic sutures, $LF+RP$ or $RF+LP$ denoting the extent of contact between the frontal and parietal bones from the sagittal to the metopic suture, and the measurement the length of the common suture between them: males—132 ($LF+RP$, 1·5), 627 ($LF+RP$, 9·7), 644 ($LF+RP$, 7·0), 700 ($LF+RP$, 10·1), 726 ($LF+RP$, 11·6), 782 ($LF+RP$, 11·3), 827 ($LF+RP$, 4·1), 895 ($LF+RP$, 5·2), 937 ($LF+RP$, 7·0), No. x ($LF+RP$, 1·0?), 1012 ($LF+RP$, 4·7), 696 ($RF+LP$, 3·2), 844 ($RF+LP$, 6·5), 860 ($RF+LP$, 5·7?), 1074 ($RF+LP$, 6·5); females—612 ($LF+RP$, 5·2), 620 ($LF+RP$, 2·9), 655 ($LF+RP$, 8·5), 744 ($LF+RP$, 3·7), 826 ($LF+RP$, 3·4), 853 ($LF+RP$, 4·2), 900 ($LF+RP$, 3·8), 1014 ($LF+RP$, 2·9), 969 ($RF+LP$, 2·1), 1050 ($RF+LP$, 5·0). All other long series examined in this way have shown that contact is most frequently made between the left frontal and right parietal bones.

The percentage frequency of occurrence of the suture is 13·4 for the males and 11·5 for the females. It has generally been found that the females are rather more affected than the males, and this was so for the Spitalfields series giving percentages of 11·3 and 8·4. Without distinguishing the sexes, Professor Parsons found 52 with "unclosed or partly closed metopic sutures" among 590 Hythe skulls, and the resulting percentage of 8·8 is apparently lower than ours owing to the fact that we have included cases where the suture was completely closed, or even partly obliterated. Regarding the state of closure of the metopic suture, this is less advanced than that of the sagittal in eight cases among the males, more advanced in two and in the same state in five; while among the females there is one case less advanced, two more advanced and seven in the same state. When the metopic suture persists to an adult stage it appears to close at about the same age as the sagittal. It has been shown for several series that some breadth measurements of metopic skulls are appreciably larger than those of non-metopic specimens on the average. Comparisons are made in the following table:

	Series	Internal bi-orbital breadth (*IOW*)	Minimum frontal breadth (*B′*)	Maximum frontal breadth (*B″*)	Maximum parietal breadth (*B*)	Biasterionic breadth
♂	Metopic skulls (A) ...	100·1 (15)	102·1 (14)	129·0 (13)	148·0 (15)	113·4 (13)
	Non-metopic skulls (B)	97·8 (95)	98·1 (95)	124·2 (89)	146·5 (97)	112·6 (84)
	A − B	+2·3	+4·0	+4·8	+1·5	+0·8
♀	Metopic skulls (A) ...	96·3 (10)	96·8 (10)	121·5 (10)	139·5 (10)	107·4 (10)
	Non-metopic skulls (B)	94·0 (76)	94·4 (77)	118·8 (74)	140·4 (77)	108·5 (77)
	A − B	+2·3	+2·4	+2·7	−0·9	−1·1

The amounts by which the means for metopic skulls exceed those for the non-metopic series arrange these breadth measurements in the same order for the two sexes and all the male differences are greater in a positive sense than the female. Precisely similar conditions were found for the Spitalfields series, though all the differences in this case are less in a positive sense than the corresponding ones for the Hythe skulls. The failure of the frontal suture to close before maturity is reached is apparently associated with an increased breadth of the frontal region as a whole, while the most marked increase is in the region of the coronal suture. Traces of the suture between the ex- and supra-occipital bones were found on two male skulls (619 *R*, 644 *R* and *L*) and three female (724 *R*, 853 *L*, 870 *L*) and the complete suture on the left side could be traced on one female specimen (867, Plate VII *c*). Another female skull (745) has the coronal and lambdoid sutures obliterated, the temporal squamae completely fused to the parietal bones and the calvarial walls falling in owing to senescence, but the anterior half of the sagittal suture is still open. No traces of horizontal sutures across the malar bones were found.

(*b*) *Supernumerary Bones.* More than 50 °/₀ of the Hythe skulls in the case of both sexes have one or more wormian bones in the lambdoid suture. As usual these are very variable in number and size. A male specimen (843) has an ossicle of bregma extending backwards from the coronal suture in the line of the sagittal suture for a length of 23·5 mm. and having a maximum breadth of 13 mm. There are also a few small ossicles in the coronal or sagittal sutures of other male skulls. There is no female specimen with an ossicle of bregma, but No. 875 has a wormian bone (15 × 7) in the left coronal suture and No. 670 has a large wormian (29 × 28) in the same position and also a supernumerary bone 27 mm. long between the left temporal squama and parietal bone above the auricular passage. There are 110 males on which a bregmatic ossicle might have been observed, so the percentage frequency is 0·9 for males, and for females it is zero. Professor Parsons found 6 cases in nearly 600 Hythe skulls, 3 belonging to either sex, and these frequencies are not unusual. Epipteric bones were counted only as such when they were found to have a maximum length of over 3 mm. There are 24 male specimens having at least one of these supernumerary bones at the pterion at the right or left, or both, sides out of 75 possible cases. Counting multiple cases, there are 20 on either side. For the females there are 17 affected out of a possible 66, there being 14 bones on the right and 17 on the left. Among the 75 male skulls having the sutures at the pterion open, there is one with fronto-temporal articulation on both sides (1070); among the 66 female skulls there is one with articulation on both sides (967) and another showing a close approach to the condition on the left (655). These frequencies appear to be quite normal for a European series. True inter-parietal bones were found in 5 out of 112 male skulls (4·5 °/₀) and 3 out of 84 female (3·6 °/₀). The male and female percentages for the Spitalfields series were respectively 1·6 and 0·5. Among an unspecified number, but probably about 550, of the Hythe skulls belonging to both sexes Professor Parsons found 13 (*ca.* 2·4 °/₀) with true inter-parietal bones, so the percentages for our smaller sample may be high owing to a fortuitous selection. Among the male specimens affected, one (1031) shows the transverse occipital suture only open, though the lambdoid suture is partly obliterated, another (894) shows the same simple inter-parietal form, one (1049) has the two *ossa triangularia* only separate, one (903) has the complete tripartite form, except that there is no suture between the *os pentagonale* and left *os triangulare*, and the last (1071) has the left *os triangulare* and the *os pentagonale* only separate, but with no suture between them. One of the affected female specimens (744) has the right *os triangulare* only separate, another (818) has the *os pentagonale* only separate and the last (875) shows the complete form except that there is no suture between the right *os triangulare* and the *os pentagonale*. Traces of the transverse occipital suture near the asteria were observed on a few other skulls. A large symmetrically placed and undivided ossicle of lambda (*os épactal*) was found on one male specimen (970) and two female (620 and 764).

(*c*) *Teeth.* Remarks in the tables of individual measurements refer to the number of teeth lost before death and to the state of wear of those remaining *in situ*, the vast majority having fallen out after death. Many are markedly worn and

only 28 showing signs of caries were found. The absence of both third molars was noted in the case of 16 male skulls (630, 648, 656, 674, 710, 726, 733, 746, 792, 827, 852, 912, 942, 1031, 1037 and 1071) and 14 female (670, 758, 762, 826, 841, 857, 867, 870, 907, 908, 911, 939, 955 and 1004). There are also four male skulls (632, 725, 732 and 1049) and four female (613, 697, 702 and 1002) with no third molar on the right side, and two female (727 and 1092) with no third molar on the left side. There seems to be no significant sexual difference between these frequencies. It is of interest to inquire whether the palate is of an unusual size when the third molars fail to develop. The means given below are respectively for the skulls with no third molar on either side and for all the others:

	Series	Palatal length (G_1)	Palatal breadth (G_2)	Palatal height (EH)
♂	No third molars ...	49·7 (14)	41·5 (12)	11·4 (9)
	Third molars erupted	51·0 (82)	41·8 (39)	12·2 (38)
♀	No third molars ...	46·8 (11)	39·3 (9)	9·5 (9)
	Third molars erupted	47·1 (53)	39·4 (35)	10·6 (33)

These figures suggest that the palate is slightly smaller on the average for the skulls having no third molars, but larger numbers would be required in order to prove that this difference is significant. A few dental anomalies were observed. A male skull (646, Plate VII *b*) has the socket for the right canine divided and what appears to be a supernumerary tooth is erupting behind the sockets of the incisors on the same side. Two female specimens (620, Plate VII *a*, and 899) have the canine erupting behind the second incisor on the left; another (955) has neither canines nor third molars erupted and another female (758) has no third molars, but there is a socket for a supernumerary tooth (lost post-mortem) between the canine and first premolar on the right side.

(*d*) *The Base of the Skull.* The relative sizes of the jugular foramina were compared and the following frequencies are found, JR or JL denoting that either the right or left side is the greater:

	JR	$J=$	JL
♂	61 (54·5 %)	15 (13·4 %)	36 (32·1 %)
♀	51 (58·6 %)	14 (16·1 %)	22 (25·3 %)

The balance in favour of the right side has been found for every other series examined in this way and very similar percentages were given by the Spitalfields skulls. As is usually observed in the case of a long series, there are various forms and sizes of precondyles present and the percentage frequency of the condition is

of no value unless it is defined in some arbitrary way. Two distinct processes were found on three male (718, 808 and 827) and five female (670, 719, 724, 818 and 864) specimens; five male (835, 869, 891, 970 and 997) have a single process, whether placed symmetrically with regard to the median sagittal plane or not, and another male (632) has two small precondyles connected by a bridge of bone. One Spitalfields skull (see Plate VII *b* in that paper) was found with a small spine projecting into the foramen posterior to the basion and there are two Hythe male (626 and 795) and two female (612 and 624) skulls showing the same condition. A female specimen (771, Plate VI *d*) has an articular surface at the anterior margin of the *foramen magnum*; another (1004, Plate VI *a*) has an irregular surface there and a *fossa pharyngea*—the only one noted—7 mm. deep, and another female (719) has a par-occipital process on the right side in the form of a blunt spine. An unusual condition is shown by a male specimen (1071, Plate VI *c*) having the left side of the basi-occipital normal, but the right considerably enlarged. A female (659, Plate VI *b*) has a clear opisthial notch.

(*e*) *Other Anomalies.* Cases of tympanic perforation were recorded and any hole in the plate through which a wire 0·28 mm. in diameter could be passed so that it entered the auricular passage was counted as such. As usual the holes were found to vary considerably in number and size. Among 112 males there are 16 affected (616, 632, 640, 644, 646, 685, 754, 792, 815, 825, 827, 863, 895, 971, 974 and 997), the total numbers of perforations being 17 on the right side and 16 on the left, and among 85 females there are 30 affected (610, 620, 636, 654, 660, 692, 702, 713, 715, 719, 724, 737, 744, 753, 758, 803, 806, 817, 834, 856, 857, 870, 939, 957, 960, 1014, 1047, 1054, 1088 and 1096), the total numbers of perforations being 36 on the right side and 45 on the left. The percentage frequency is thus 14·3 for the males and 35·3 for the females, while a negro series from Kenya Colony examined in precisely the same way* gave 46·1 % for the males and 53·3 % for the females. This condition appears to exhibit distinct racial and sexual differences. Although it is generally supposed to be influenced by the age of the individual, this supposition was not confirmed in the case of the negro series referred to. The table below gives the percentage with perforated tympanic plates occurring in each of the age groups distinguished from the condition of the principal sutures, the totals for these groups having been given in Table III above:

	Sutures open	Sutures beginning to close or partly closed	All sutures closed
♂	13·0 %	14·5 %	16·0 %
♀	41·2 %	34·6 %	0·0 %

* Elisabeth Kitson: "A Study of the Negro Skull with Special Reference to the Crania from Kenya Colony," *Biometrika*, Vol. XXIII. 1931, pp. 271—314.

These figures again suggest that there is little, if any, association between the age of the individual at death and the condition in question: it is certainly possessed by some young adults and some aged persons are without it. A male specimen (895, Plate VII *d*) has a large perforation, having roughly the form of a circle 0·8 mm. in diameter, on the left side and the outer border of the element is formed by a thin and irregular bridge of bone. The articular surface of the glenoid fossa on the same side is enlarged and roughened as the result of arthritis. This surface on the right is only slightly roughened and the tympanic plate is normal. A female specimen (1054, Plate VII *e* and *f*) has quite anomalous auricular passages and the forms on the two sides are very similar. Each tympanic plate has six perforations and thin and very irregular outer edges. The posterior walls of the passages are also irregular and defe tive in parts. Such an irregularity appears to be quite rare and it may be remembered that a male Spitalfields skull showed the complete absence of the posterior wall of the right auricular passage (see Plate VI *d* in that paper). The only exostoses having their maximum length more than 6 mm. were one on the left parietal bone of a male specimen (941), one on the occipital bone near the right mastoid process on a female (900) and one on the left parietal of a male (No. *x*). The last is the largest; it has a circular base with a diameter of 35 mm. and a height of 11 mm. There are no marked examples of a post-coronal depression, but a peculiarity which may be called a post-coronal eminence was noted in the case of two male (663 and 929) and three female (753, 806 and 828) specimens. The most marked of these (929) is shown in Plate V *a*. There is a distinct eminence not immediately behind the coronal suture, but nearer the normal position of the vertex. There are a few examples of a low median sagittal ridge and these occur more frequently among the metopic than among the non-metopic skulls. A few specimens have unusually retreating frontal bones, but these are exceptional cases, and again a few have an asymmetrical calvaria, though this condition is never exaggerated. Two female skulls (875 and 967, Plate V *b*) have an unusually globular form of calvaria, but these appear to be extremes of normal variation in that direction rather than examples of hydrocephaly. The only distinct traces of disease, apart from dental caries, are the arthritic glenoid surfaces on two crania, one male (895, Plate VII *d*) and one female (1010). Professor Parsons says: "Ten or twelve examples of syphilitic lesions were found in the form of ulcerations, necroses, gummata and periostitis. One skull with a large gummatous heaping up of bone has been shown to countless visitors in the past as an instance of a healing wound." The evidence for the occurrence of syphilis appears to have been derived principally from the long-bones.

(6) *Comparison between the Hythe and Spitalfields Series.* In the paper dealing with the Spitalfields series it has been shown that there is a peculiarly close resemblance between that type and the one represented by the Hythe skulls. Precisely similar methods of technique have been used in describing the two samples and it will be profitable to make a somewhat detailed comparison between them. The frequencies with which anomalous conditions occur can give no definite

measure of the racial affinities of two series in the present state of our knowledge and large numbers would be needed in order to give reliable determinations. As far as can be seen, the two English series are very similar to one another in these respects. The frequency of occurrence of the metopic suture is unusually high for both and both have examples of malformed dentitions and auricular passages.

Using coefficients of variation of the absolute measurements and standard deviations of the indices and angles, the variability of our Hythe sample is compared in the table below with that of the Spitalfields and with those of two other European series:

Series		No. of characters with greater variabilities than Hythe	No. of characters with variabilities equal to Hythe	No. of characters with lesser variabilities than Hythe	No. of characters compared
Farringdon St. English (17th century)	♂	46 (68·7 %)	1 (1·5 %)	20 (29·9 %)	67
	♀	44 (65·7 %)	0 (0·0 %)	23 (34·3 %)	67
Spitalfields	♂	46 (63·0 %)	2 (2·7 %)	25 (34·2 %)	73
	♀	18 (50·0 %)	2 (5·6 %)	16 (44·4 %)	36
Basque	♂	19 (57·6 %)	0 (0·0 %)	14 (42·4 %)	33

Comparisons made in this way are far from exact since the number of characters used is not constant and different measurements of the same feature are included in them, such as the different nasal heights or orbital breadths, but there is a clear suggestion that the Hythe series is less variable than these other three. Similar comparisons of English series and the Basque have been made previously and their relative variabilities give the decreasing order: Farringdon St.—Whitechapel—Spitalfields—Basque—Hythe. The differences between adjacent series according to this grading are certainly very small and they are, perhaps, not larger than those which might arise from random sampling. It may be noticed that the evidence for racial homogeneity furnished by a considerable number of characters, as in the present instance, may be appreciably different from that which a single measurement suggests. The standard deviations of the male cephalic indices for the five series compared above and in the order given there are: $3{\cdot}48 \pm {\cdot}14$, $3{\cdot}26 \pm {\cdot}14$, $3{\cdot}84 \pm {\cdot}11$, $2{\cdot}68 \pm {\cdot}15$ and $3{\cdot}69 \pm {\cdot}17$. If this character had been considered alone, our conclusions would hence have been substantially different and it is evident that the only safe method of comparing the relative variabilities of several series is one which takes into account the constants relating to a number of different measurements. The samples with which we are dealing are generally small ones for statistical purposes and it is not surprising to find that the majority of the differences observed are insignificant. Of the 73 characters for which constants of variability can be compared in the case of the Hythe and Spitalfields male series, there are only four for which the difference exceeds three times its

probable error. These are: SS (Δ/p.e.Δ = 3·1), 100 $NB/NH, L$ (3·5), GB (4·4) and 100 G_2/G_1' (3·2). The Hythe constant is the less in the first two cases and the greater in the others. For the female series there are only 36 comparisons possible, since constants of variability were not calculated for fewer than 30 skulls, and only one difference is significant, viz. fmb (3·5), the Spitalfields constant being the greater. It was found that the Spitalfields male coefficient of variation for GB was an outstandingly low value (3·98 ± ·21); the Hythe value (5·45 ± ·26) is close to several others available and there are no variabilities for the latter series which are extreme.

The Hythe male and female means are given in Table II. The following lists give the ratio of the difference from the corresponding Spitalfields mean to the probable error of the difference for all out of the possible 77 cases having this ratio greater than three:

Male—C (3·3), F (4·1), F.V. L (5·7), L (6·1), B (6·5), B'' (4·2), Biasterionic B (4·8), H (4·6), S_2' (5·7), S_2 (4·2), Q' (4·7), Bregmatic Q' (3·4), Broca's Q' (3·6), fml (3·1), fmb (3·6), J (5·1), NH, R (3·7), NH, L (4·3), NH' (3·1), DC (4·3), DA (5·9), G_1 (4·3), G_1' (5·1), 100 B/L (11·1), 100 H'/L (6·4), 100 H/L (7·2), 100 B/H' (3·4), 100 $(B-H')/L$ (3·7), 100 $NB/NH, R$ (3·1), 100 $NB/NH, L$ (3·4), 100 NB/NH' (3·1), 100 $O_2/O_1, R$ (3·7), 100 O_2/Lacr. O_1 (3·0).

Female—C (4·2), B (5·6), B'' (3·4), Biasterionic B (4·0), H (4·3), U (3·6), Broca's Q' (3·1), fmb (5·5), DA (3·8), SS (3·1), G_2 (3·1), O_1, R (5·7), O_1, L (5·5), O_2, R (4·3), O_2, L (3·1), 100 B/L (5·4), 100 B/H' (3·1), 100 $(B-H')/L$ (3·4), 100 fmb/fml (3·2).

There are 33 characters showing a significant difference between the male means and only 19 between the female. This discrepancy is merely due to the fact that the male means are based on larger numbers of individuals than the female, and it will be shown below that there is no reason to suppose that the types for one sex resemble one another more closely than those for the other. In comparisons of this kind it is not at all unusual to find that the difference between two means exceeds 50 or 100 times its probable error in the case of the most divergent characters, and the fact that the largest ratio above is as low as 11·1 indicates that the types are very similar. If samples of 10—20 individuals had been the only material available for one sex, it is likely enough that no single character could have been proved to indicate a significant difference. Between the male means the highest ratio considered is for the cephalic index (11·1), and the only other values greater than 6 are for the height-length indices and the calvarial maximum length and breadth. Measurements of the calvaria are able to differentiate the types more effectively than those of the face. The following conclusions may be deduced from a comparison of the two series of means, the more reliable male constants being considered more than the female. The Hythe type has the lesser length, but the greater breadth and height of the brain-box and its capacity and transverse arc are hence the greater. It is of interest to note that the maximum calvarial breadth shows a more significant difference than any other breadth measurement, though

the bizygomatic differentiates the types nearly as well. The difference between the total calvarial lengths is associated with that between the median sagittal lengths of the parietal bones from bregma to lambda, but the lengths of the frontal and occipital bones in the same plane cannot be distinguished. The Hythe type has the greater length and breadth of the *foramen magnum*. Significant differences between the calvarial indices involving the length, breadth and height are to be expected from these relations and the Hythe male cephalic index (82·6 ± ·24 (112)) is three points greater than the Spitalfields (79·4 ± ·16 (274)). The only clear distinctions between the facial characters are those associated with the greater nasal height and hence lower nasal index for the Hythe series and again with its lesser dacryal arc and chord. There is also a suggestion that the Hythe type has the greater orbital height, palatal length and orbital index. It should be realised that the absolute differences between the Hythe and Spitalfields means are in all cases very small. The calvarial breadths show a more significant difference than any other chords, but the Hythe male mean only exceeds the other by 2·6 mm. which is 1·8 °/₀ of the measurement. If two skulls were available possessing all the mean measurements of the two series it is unlikely that any distinction between their facial skeletons could be appreciated from a visual examination. The only clear difference between the brain-boxes would be in the case of the character expressed by the cephalic index, the maximum lengths and breadths on which it depends only differing slightly but still in opposite senses, so that there is an appreciable difference in the form of the outline seen in *norma verticalis*.

A comparison of the Spitalfields mean measurements with those available for other European series showed that the London population was characterised by an extremely small nasal height and an extremely high nasal and low orbital index. For these three characters the Hythe are close to the Spitalfields values and nearer than they are to the inter-racial mean. The London type was also found to have a shallower and rounder palate (judging by the indices 100 EH/G_2 and 100 G_2/G_1) and a flatter nasal bridge (judging by the index 100 DS/DC) than any other European race, though the comparative material for these characters is very inadequate. Neither in these ways, nor in any other which can be estimated from the measurements, is the Hythe type peculiar, and most of its characters are typically European.

Comparisons by the method of the coefficient of racial likeness may now be considered. The Farringdon St. standard deviations were used for this purpose, since several hundred coefficients have previously been calculated with them, and slightly lower values are hence to be expected than if the constants for the more homogeneous Spitalfields or Hythe series were used. Between these two a crude male coefficient of 5·18 ± ·17 is found for 31 characters. The Spitalfields $\bar{n}$ is 163·6 and the Hythe 101·3, giving a reduced coefficient of 4·14 ± ·14. The female palatal breadth and index were omitted in calculating the female values since the Spitalfields means are based on fewer than 10 crania. For the remaining 29 characters the crude coefficient is 3·08 ± ·18. The Spitalfields $\bar{n}$ is 50·6 and the Hythe 81·1,

giving a reduced coefficient of $4{\cdot}93 \pm {\cdot}28$. The difference between the male and female crude coefficients is markedly significant, and this was to be expected since the numbers of crania for the two sexes differ markedly. Correction is made for this difference by reducing the constants, and the difference of 0·79 found after this has been done is only 2·5 times its probable error. In the case of each sample we have every reason to believe that the male and female series represent precisely the same population, and the absolute racial divergencies between the male means and between the female means should hence be of the same order. The male means will give the better measure of this divergence since they represent larger numbers than the female. The α's found in computing the coefficients give another measure of the significance of a difference in the case of each character compared. The only values greater than 10 in the case of the males are: $100\,B/L$ ($\alpha = 67{\cdot}27$), $100\,H'/L$ (19·67), B (16·45), L (15·78) and J (11·09); and in the case of the females $100\,B/L$ (22·13), O_1, R (16·95), B (16·11) and *fmb* (14·15). The outstanding difference is the one between the cephalic indices.

A final comparison may be made between the type contours of the two very similar series. The mean measurements used in constructing the Hythe figures are given in Tables IV—VI and the types are Figs. 3—8. The Hythe contours for both sexes are based on adequate numbers of crania, but there were only 24 female Spitalfields skulls which could be used for this purpose. The tracings provided greatly facilitate the comparisons. Superposing the transverse sections, it is found that the Spitalfields outlines fall entirely inside the Hythe. The difference between the heights is very small in the case of the male figures and the maximum divergencies are found between the 4th and 5th parallels where both types reach their maximum breadths. The maximum distance between the outlines here is 2·0 mm. on the right side and 3·2 on the left. Rather larger divergencies are found between the female sections. The horizontal sections are most conveniently superposed by making the points F coincide and the FO axes covering one another. The Spitalfields male outline then falls inside the Hythe for three-quarters of its length, but the two cross between the 9th and 10th parallels and there is a difference in length of 2·9 mm. The maximum breadths of both types are, as usual, between the 6th and 7th parallels, and it is in this region that the maximum divergencies are found if the sections are moved until the two F's and the two O's are 1·45 mm. from each other. The maximum divergence is then 2·1 mm. on the right side and 2·0 mm. on the left. The female Spitalfields type falls entirely within the Hythe, the difference between the lengths being 1·0 mm., and the maximum divergencies between the outlines fall between the 6th and 8th parallels. It is clear from the comparison of these transverse and horizontal sections that the greatest absolute difference between the brain-boxes of the types occurs in the region of their maximum parietal breadths, and this has been shown to be the absolute measurement which distinguishes the male series more effectively than any other. As may be shown from other comparisons, a difference in this region may be larger and more significant than those between all frontal or occipital breadths.

TABLE IV.

Hythe Transverse Vertical Contours. Mean Values.

	MA	1*R*=1*L*	2*R*	2*L*	3*R*	3*L*	4*R*	4*L*	5*R*	5*L*	6*R*	6*L*	7*R*
♂	114·9 (112)	59·9 (112)	65·6 (111)	65·6 (112)	69·9 (111)	69·9 (112)	72·4 (108)	72·3 (108)	72·3 (107)	72·6 (106)	71·1 (111)	71·7 (112)	68·9 (112)
♀	110·4 (86)	57·6 (86)	63·4 (85)	62·9 (86)	67·4 (83)	66·6 (86)	69·5 (83)	68·8 (86)	69·3 (85)	68·8 (86)	68·5 (86)	67·9 (86)	66·7 (86)

	7*L*	8*R*	8*L*	9*R*	9*L*	10*R*	10*L*	*A*½*R*	*A*½*L*	*ZR, R*		*ZR, L*	
										x	*y*	*x*	*y*
♂	69·5 (112)	64·1 (112)	64·7 (112)	54·5 (112)	55·1 (112)	38·0 (112)	39·5 (112)	17·2 (112)	19·4 (112)	2·2 (112)	63·5 (112)	2·4 (112)	63·5 (112)
♀	66·3 (86)	62·0 (86)	62·1 (86)	53·0 (86)	53·1 (86)	37·4 (81)	37·5 (86)	17·5 (86)	17·9 (86)	2·7 (86)	61·0 (86)	2·7 (86)	60·8 (86)

TABLE V.

Hythe Horizontal Contours. Mean Values.

	FO	*F*¼*R*	*F*¼*L*	*F*½*R*	*F*½*L*	2*R*	2*L*	2½*R*	2½*L*	3*R*	3*L*
♂	177·1 (112)	24·9 (112)	23·7 (112)	35·4 (111)	34·5 (112)	47·8 (111)	47·4 (112)	50·5 (112)	50·2 (112)	52·9 (112)	52·3 (112)
♀	170·7 (87)	22·5 (87)	22·8 (87)	34·1 (87)	35·1 (87)	46·0 (87)	47·2 (87)	48·8 (87)	48·8 (87)	50·9 (85)	51·2 (87)

	4*R*	4*L*	5*R*	5*L*	6*R*	6*L*	7*R*	7*L*	8*R*	8*L*
♂	59·3 (112)	58·7 (112)	67·0 (110)	66·1 (111)	72·1 (111)	71·1 (112)	73·6 (112)	72·4 (112)	70·6 (112)	69·4 (112)
♀	56·0 (84)	57·2 (87)	63·0 (83)	64·4 (87)	68·1 (85)	69·1 (87)	69·3 (87)	70·0 (87)	66·3 (87)	66·9 (87)

	9*R*	9*L*	10*R*	10*L*	*O*½*R*	*O*½*L*	*T*(*R*)*y*	*T*(*L*)*y*	*T*(*R*)*x*	*T*(*L*)*x*
♂	62·2 (112)	61·5 (111)	47·0 (112)	46·6 (111)	25·9 (112)	25·8 (112)	49·9 (111)	49·4 (111)	19·7 (111)	19·7 (111)
♀	58·2 (87)	58·7 (87)	43·4 (87)	44·3 (87)	24·3 (87)	25·2 (87)	48·1 (87)	48·0 (87)	19·2 (87)	17·7 (87)

TABLE VI.

Hythe Sagittal Contours. Mean Values.

	$N\gamma$	Ordinates above $N\gamma$												
		$O=N$	$N\frac{1}{4}$	1	2	3	4	5	6	7	8	9	$\gamma\frac{1}{4}$	$\gamma\frac{1}{8}$
♂	174·7 (112)	22·1 (112)	36·8 (112)	58·3 (112)	71·4 (112)	79·1 (112)	83·6 (112)	85·5 (112)	85·8 (112)	82·6 (112)	74·4 (112)	54·1 (112)	27·2 (112)	20·2 (112)
♀	168·6 (87)	24·6 (87)	40·2 (87)	58·0 (87)	69·5 (87)	76·5 (87)	80·9 (87)	83·0 (86)	83·4 (86)	80·3 (87)	72·3 (87)	53·8 (87)	28·0 (87)	20·8 (87)

	Ordinates below $N\gamma$							Vertex		β		Glabella	
	$N\frac{1}{4}$	1	2	8	9	$\gamma\frac{1}{2}$	$\gamma\frac{1}{4}$	x from N	y	x from N	y	x from N	y
♂	63·6 (103)	57·3 (111)	54·5 (110)	49·1 (112)	39·5 (112)	29·9 (112)	21·5 (112)	95·9 (112)	86·4 (112)	73·2 (112)	84·0 (112)	2·9 (112)	8·6 (112)
♀	58·5 (78)	52·6 (87)	50·5 (84)	46·7 (87)	35·8 (87)	26·0 (87)	17·9 (87)	94·1 (86)	84·1 (86)	69·1 (87)	81·3 (87)	2·3 (87)	10·7 (87)

	Occipital Pt.		λ		Sub-Orb.		Aur. Pt.		Opisthion		Inion		Basion	
	x from γ	y^*	x from γ	y	x from N	y	x from γ	y	x from γ	y	x from γ	y	from γ	from N
♂	0·4 (112)	2·0 (112)	6·6 (112)	31·0 (112)	9·3 (112)	29·1 (112)	87·7 (112)	29·0 (112)	51·4 (112)	53·0 (112)	9·7 (111)	32·7 (111)	99·8 (112)	100·6 (112)
♀	0·5 (87)	3·1 (87)	6·1 (87)	32·0 (87)	8·9 (87)	27·8 (87)	84·8 (87)	27·9 (87)	49·3 (86)	51·0 (86)	11·1 (87)	30·7 (87)	95·6 (86)	96·3 (86)

	Alv. Pt.		Nose					Palate				Frontal	
								P'		P		Max. Sub. to $N\beta$	
	from N	from Bas.	(i)	(ii)	(iii)	$\angle LN\gamma$	NL	x from N	y	x from Alv.	y	x from N	y
♂	70·0 (91)	94·5 (91)	1·3 (107)	3·2 (106)	6·6 (88)	121°·4 (37)	21·9 (37)	47·2 (108)	52·1 (108)	35·9 (91)	15·2 (91)	50·6 (112)	25·6 (112)
♀	65·2 (64)	91·4 (64)	1·1 (82)	3·1 (74)	7·0 (59)	120°·9 (20)	21·9 (20)	44·3 (78)	48·1 (78)	35·2 (63)	14·2 (63)	47·4 (87)	25·6 (87)

	Occipital Max. Sub. to λ Op.		Max. Sub. to $N\lambda$		Max. Sub. to GI		Sp.		Sub. from $\frac{1}{2}$ Bas.—Sp. Chord		N. S.		Vert. Tang. to Alv.—N. S. Arc	
	x from λ	y	x from N	y	x from G	y	x from N	y	x from Bas.	y	x from N	y	x from N	y
♂	53·3 (112)	27·7 (112)	84·3 (112)	69·8 (112)	98·2 (111)	102·2 (111)	69·1 (112)	33·5 (112)	12·4 (112)	0·2 (112)	6·2 (94)	50·1 (94)	4·1 (97)	55·3 (97)
♀	50·0 (86)	25·4 (86)	81·6 (87)	67·0 (87)	95·7 (87)	98·1 (87)	66·2 (86)	32·0 (86)	11·6 (85)	0·5 (85)	5·8 (82)	47·2 (82)	4·1 (75)	52·4 (75)

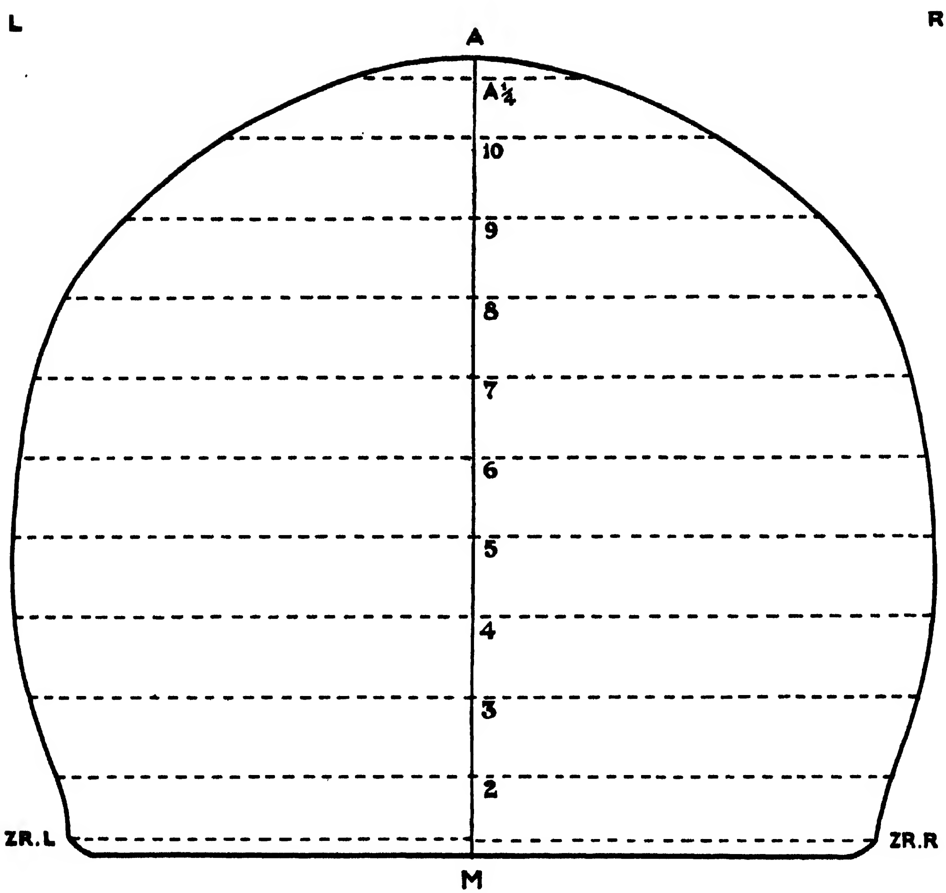


Fig. III Transverse Type Contour of 112 ♂ Hythe Skulls.

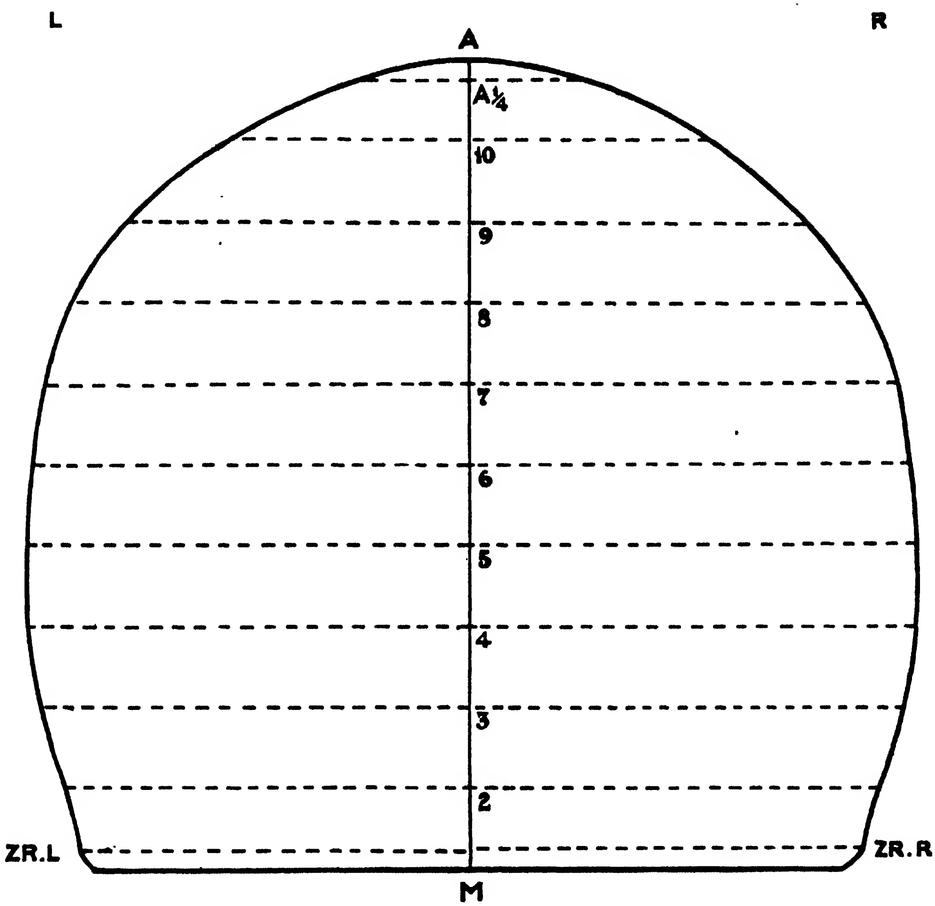


Fig. IV Transverse Type Contour of 86 ♀ Hythe Skulls.

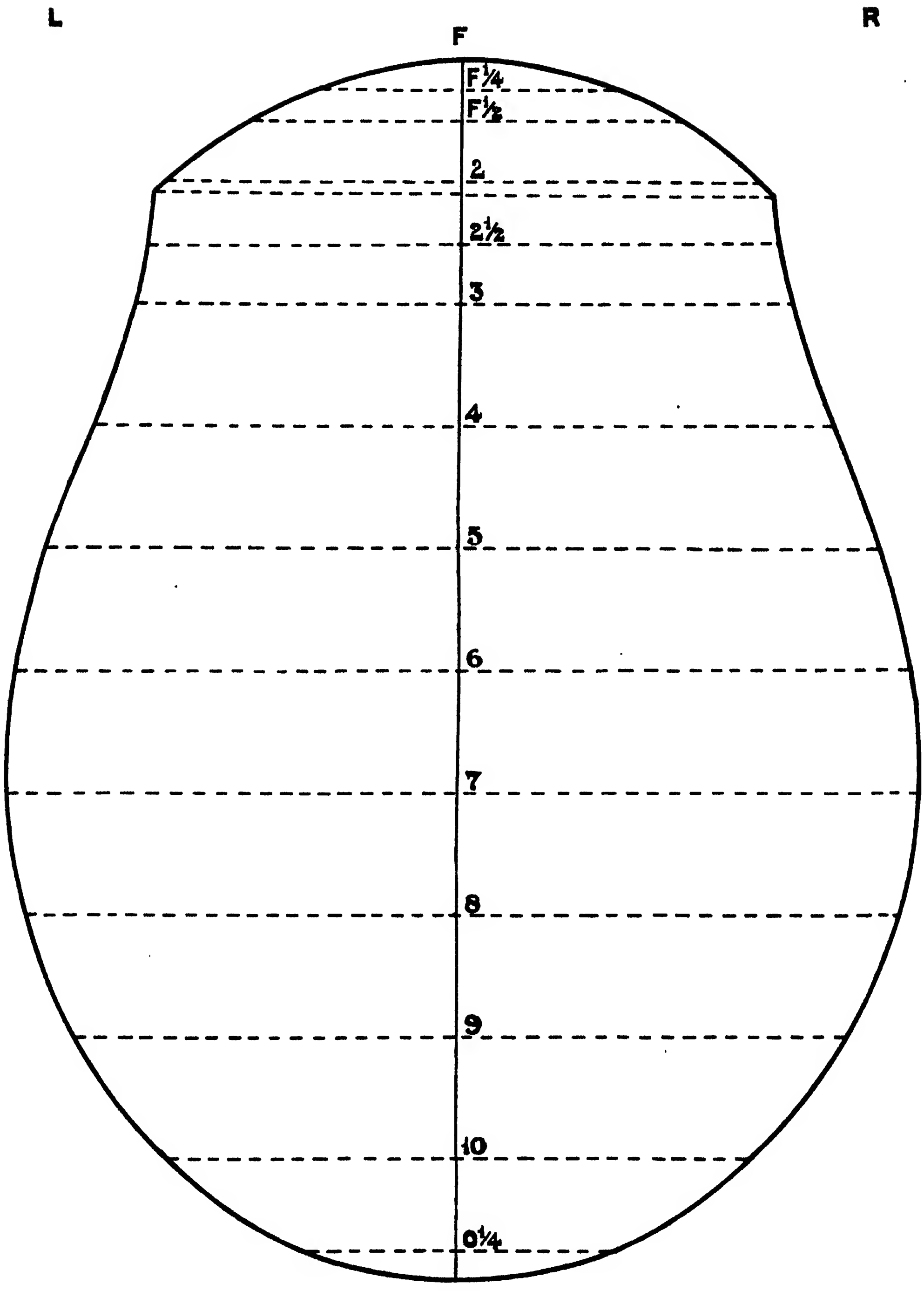


FIG. V Horizontal Type Contour of 112 ♂ Hythe Skulls.

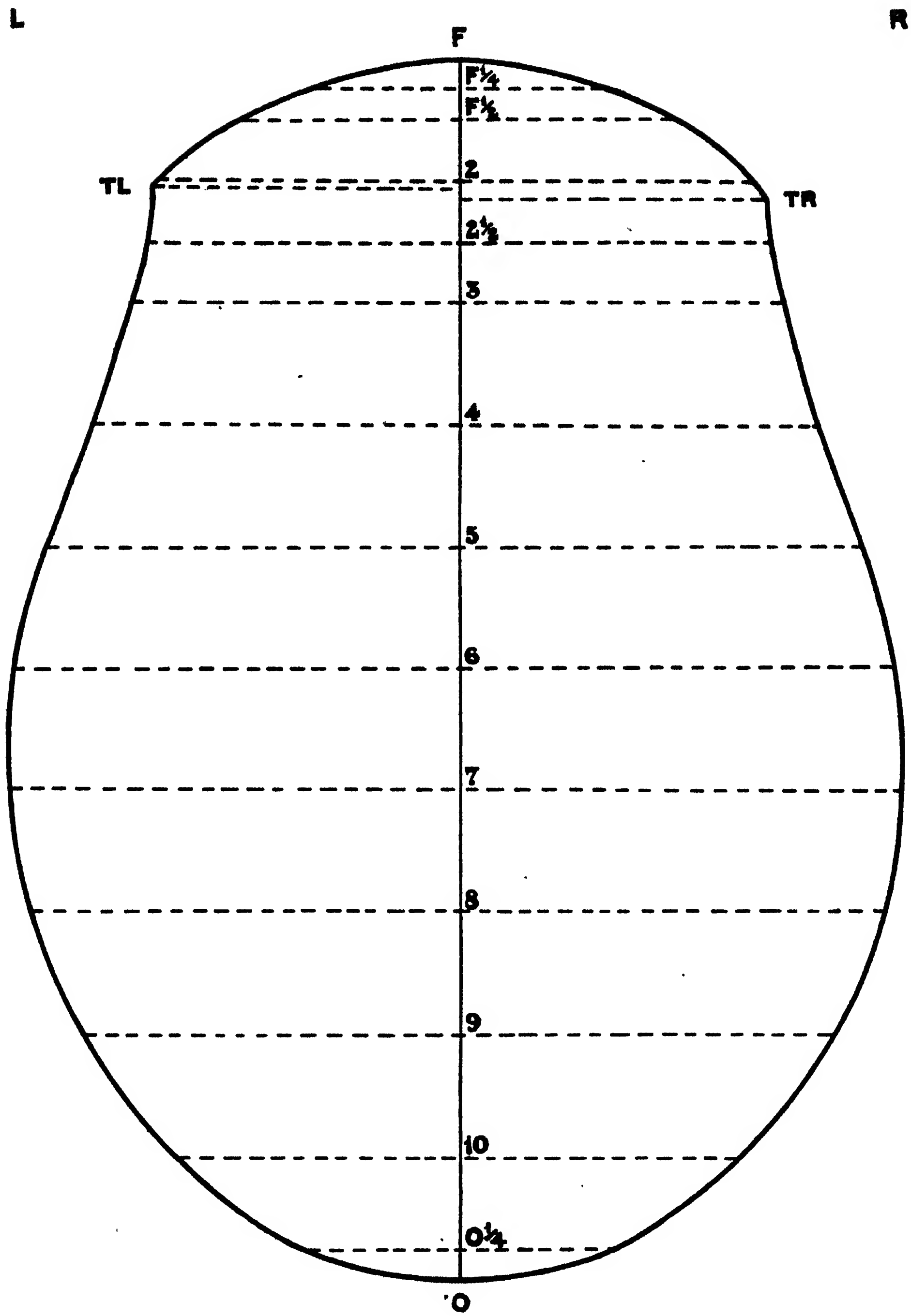


Fig. VI Horizontal Type Contour of 86 ♀ Hythe Skulls.

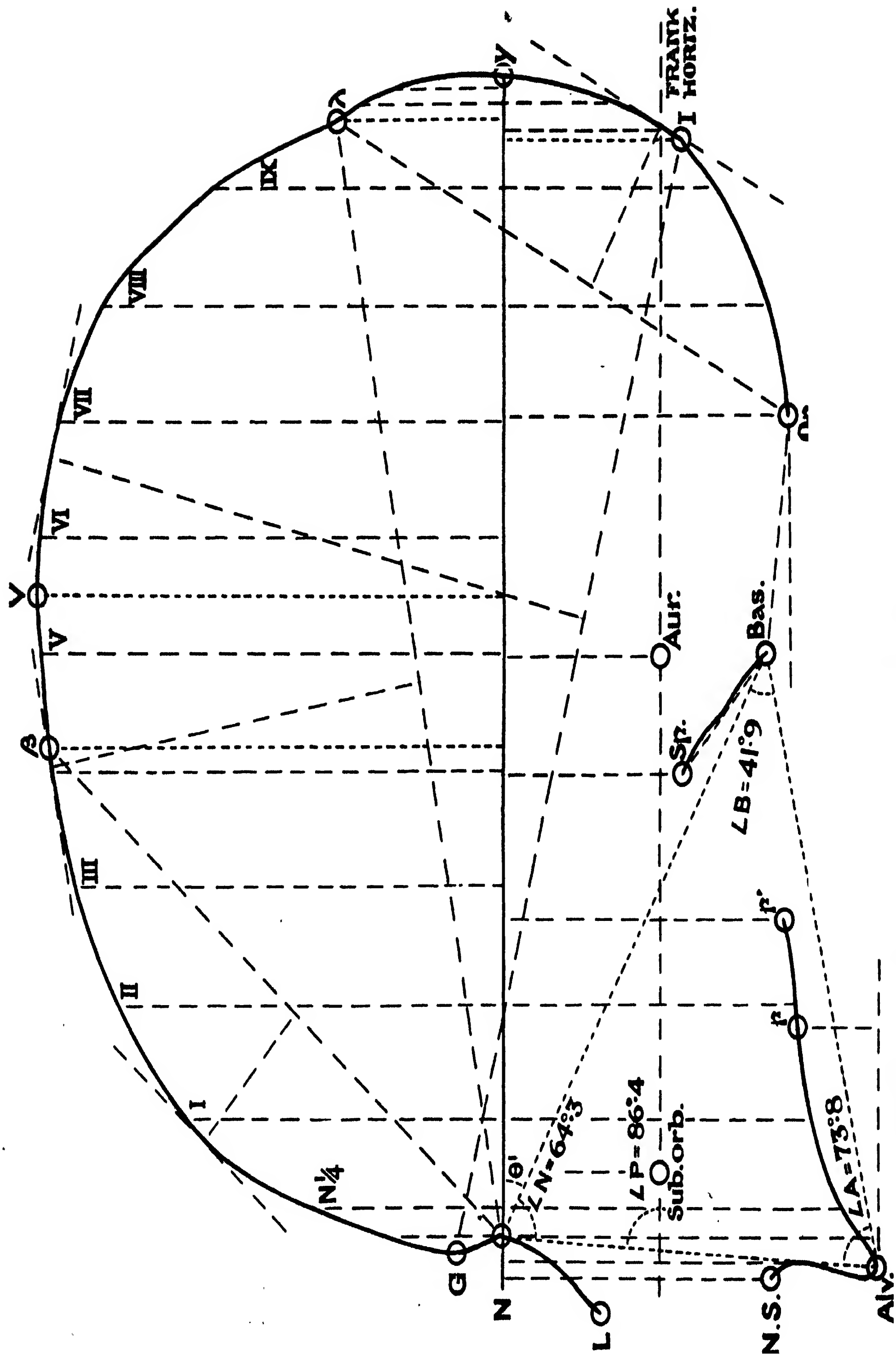


Fig. VIII Sagittal Type Contour of 112 ♂ Hythe Skulls.

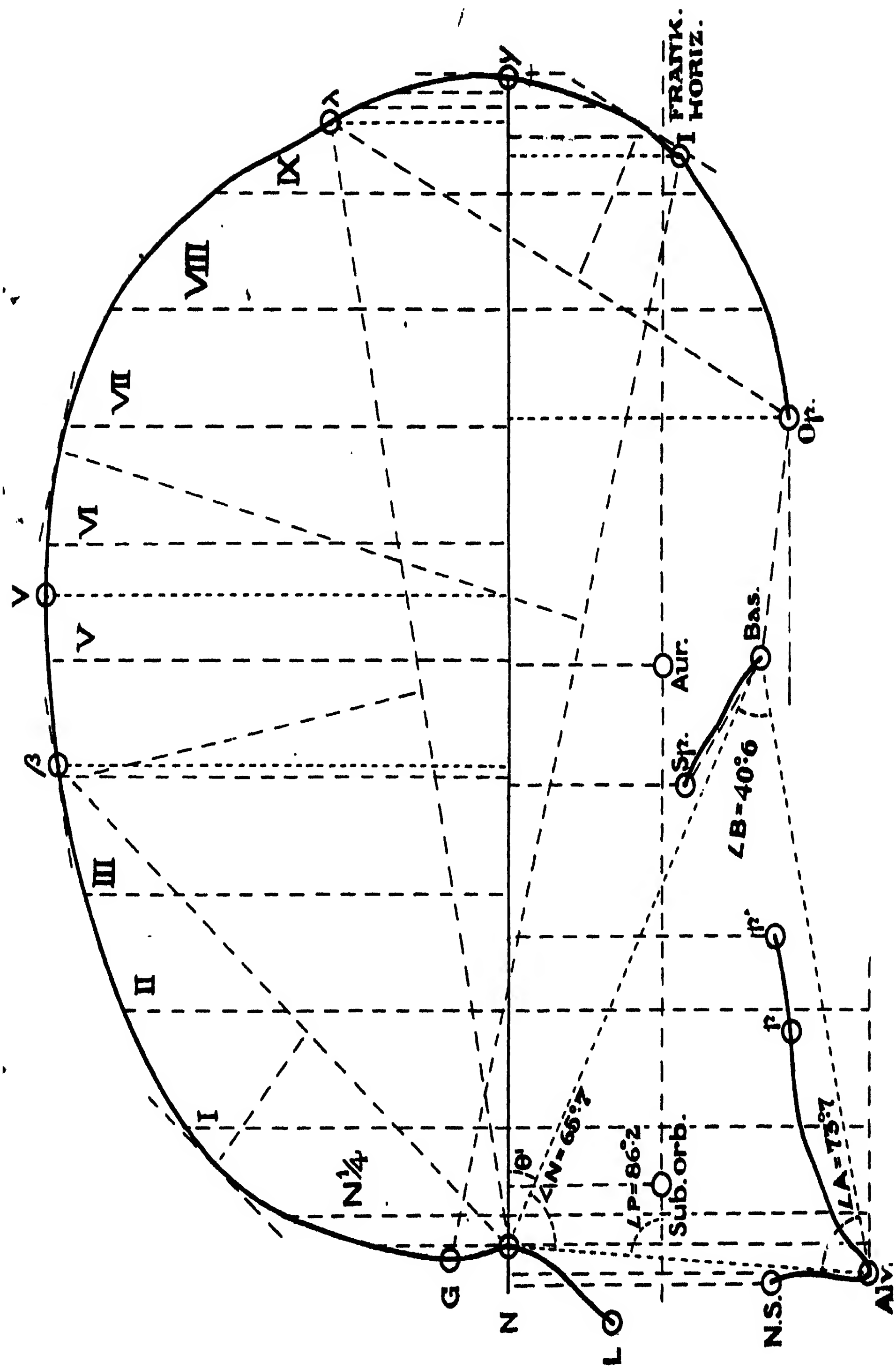


Fig. VIII Sagittal Type Contour of 87 ♀ Hythe Skulls.

The sagittal sections are most conveniently superposed by making the points N (the nasions) coincident and the $N\gamma$ axes covering one another. When this is done there is found to be a remarkably close similarity between the two types. The difference between the length is again apparent, the Spitalfields $N\gamma$ line being 2·9 mm. longer than the Hythe. The outlines of the nasal and frontal bones are practically coincident and the vertices and vaults also overlap. Behind the eighth parallel the Spitalfields outline gradually diverges beyond the other until the maximum difference is reached near the γ, then below the base line the outlines converge to meet at the inion and the arcs from that point to the opisthion cover one another. The basi-occipital surfaces between the basions and spheno-basions are practically coincident and the outlines of the palate and pre-maxillae almost exactly so. The Spitalfields auricular point is 3·7 mm. behind the Hythe*. By rotating one outline until the bases of the occipital bones—the λ *Op.* lines—coincide it can be seen that the two outlines of those bones are practically identical. It is a most significant fact that the greatest divergence disclosed in this way is one between the maximum lengths of the calvariae. This is associated with a difference between the sagittal lengths of the parietal bones and between the positions of the auricular points, but at the same time the sizes and shapes of the sections of the facial skeleton, of the base of the skull and of the frontal and occipital bones are undifferentiated. On superposing the female sagittal contours it is found again that there is a very close correspondence between the outlines of the facial skeleton and palate and of the base of the skull. The calvarial sections do not, however, show the same relationship as before. The outlines of the frontal bones diverge near the second parallel and the Spitalfields section falls inside the Hythe until they approach again near the opisthion. The fact that the Spitalfields sagittal contour is larger for the one sex and smaller for the other may be supposed due to the inadequate size of the sample on which its female type is based.

The foregoing comparison between the two series from London and Kent has shown how closely similar they are. Judging from the more reliable male samples, the most essential differences between the types are in the maximum calvarial length (the Spitalfields being the greater) and in the maximum calvarial breadth (the Hythe being the greater). The only other sagittal length differentiated is that of the parietal bones from bregma to lambda, while the bi-zygomatic diameter distinguishes the types more effectively than any other breadth except the bi-parietal. The fact that the maximum length and breadth differ in opposite senses leads to a very appreciable difference between the cephalic indices. Other calvarial indices involving the length and breadth also make a clear, though less pronounced, distinction between the types, but the facial skeletons of the two appear to be almost identical in all respects. These relationships suggest forcibly that the differences between the two varieties of cranium may have been occasioned by a single factor. An increase in the maximum length or breadth of the brain may

* An error was made in recording the x-co-ordinates of the auricular points for the male and female Spitalfields sagittal contours. The measurements given (Table XIII in that paper) are from N and not, as is stated, from γ.

be supposed to be compensated for by a decrease in the other major diameter, so that there is little change in the volume of the organ. The skull, in this particular comparison, appears to have been affected principally in the parietal and temporal regions. The difference in length is accounted for almost entirely by the difference in the median sagittal lengths of the parietal bones and the difference in breadth is most marked near the centres of those bones (where the maximum breadth falls), but there is also an appreciable change in the widths of the occipital, frontal and temporal extremities of the parietal regions. Apart from this slight difference in breadth, the occipital and frontal bones, and the facial skeleton, appear not to have been modified in either size or shape.

(7) *Comparisons with British Cranial Series.* Coefficients of racial likeness between the Hythe and eight other British cranial series are given in Table VII; female comparisons can only be made in six cases. The values based on all available of the total 31 characters are supposed to give the best measure of racial relationship. These reduced values show clearly that the Hythe stands far closer to the Spitalfields than to any other English type. The coefficients between the Spitalfields series and the others have been given, and by comparing these two lists we may conclude that the two most recently described English types are closely allied to one another and widely removed from all the others. The English Bronze Age type resembles the Hythe rather more closely than it does the Spitalfields, but it is the latter which is nearer to the seventeenth century Londoners, Anglo-Saxons, Iron Age and Neolithic peoples. It is unsafe, however, to attach any significance to different degrees of distant relationship. When indices and angles are considered alone, the Hythe coefficients with the Spitalfields and English Bronze Age types are found to be of the same order. The shapes of these three are quite similar, but the Bronze Age skull is very significantly larger than the others. As is usually found, the facial characters are less capable of discriminating the types than are the calvarial ones. This is most noticeably so in the case of the comparisons with the Moorfields series. The reduced coefficient for facial characters is the second lowest male value of this kind in Table VII and the female value is the lowest of all and actually insignificant. For the same group of characters the Moorfields was found to have the lowest male reduced coefficient with the Spitalfields series and the female value was also insignificant*. This is a curious result,

* The coefficients of racial likeness between the Spitalfields and other British series are given in Table IV of the Spitalfields paper. An error was made in computing the female coefficients with the Moorfields series. The values below are the corrected ones, the $\bar{n}$'s and coefficients for facial characters remaining unchanged.

Crude Coefficients			Reduced Coefficients		
All Characters	Indices and Angles	Calvarial Characters	All Characters	Indices and Angles	Calvarial Characters
15·96 ± ·18 (29)	17·90 ± ·29 (11)	29·00 ± ·24 (16)	36·25 ± ·41	44·25 ± ·72	48·42 ± ·40

TABLE VII.

Coefficients of Racial Likeness between the Hythe and other British Series.*

	Series	Crude Coefficients				Reduced Coefficients			
		All Characters	Indices and Angles	Calvarial Characters	Facial Characters	All Characters	Indices and Angles	Calvarial Characters	Facial Characters
Male	Spitalfields ...	5·18 ± ·17 (31) [101·3, 163·6]	7·59 ± ·28 (12) [96·3, 152·1]	8·40 ± ·24 (16) [111·4, 236·7]	1·75 ± ·25 (15) [90·5, 85·7]	4·14 ± ·14	6·43 ± ·23	5·54 ± ·16	1·99 ± ·28
	English Bronze Age	7·23 ± ·19 (26) [103·4, 29·6]	1·79 ± ·32 (9) [101·0, 19·6]	10·07 ± ·25 (14) [111·4, 37·7]	3·91 ± ·28 (12) [94·0, 20·0]	15·71 ± ·41	5·45 ± ·97	17·88 ± ·45	11·85 ± ·83
	Whitechapel ...	32·21 ± ·17 (30) [100·9, 91·0]	48·65 ± ·29 (11) [94·8, 82·6]	58·07 ± ·25 (15) [111·3, 120·2]	6·35 ± ·25 (15) [90·5, 61·7]	33·66 ± ·18	55·10 ± ·33	50·24 ± ·21	8·65 ± ·34
	Farringdon Street	33·39 ± ·17 (31) [101·3, 95·1]	48·00 ± ·28 (12) [96·3, 89·0]	58·55 ± ·24 (16) [111·4, 119·1]	6·55 ± ·25 (15) [90·5, 69·5]	34·04 ± ·17	50·89 ± ·30	50·87 ± ·21	8·33 ± ·31
	Moorfields	15·09 ± ·17 (30) [102·3, 27·5]	20·60 ± ·28 (12) [96·3, 24·3]	26·93 ± ·24 (16) [111·4, 36·7]	1·56 ± ·25 (14) [91·9, 16·9]	34·82 ± ·40	53·09 ± ·71	48·77 ± ·43	5·46 ± ·89
	British Iron Age ...	25·44 ± ·21 (20) [105·8, 54·7]	51·26 ± ·39 (6) [101·5, 52·0]	36·74 ± ·28 (12) [111·7, 54·6]	8·50 ± ·34 (8) [97·0, 54·9]	35·27 ± ·30	66·78 ± ·51	50·10 ± ·37	12·12 ± ·48
	Anglo-Saxon ...	20·70 ± ·17 (30) [101·0, 35·5]	25·84 ± ·28 (12) [96·3, 31·4]	36·12 ± ·25 (15) [111·5, 44·7]	5·29 ± ·25 (15) [90·5, 26·2]	39·40 ± ·33	54·55 ± ·58	56·60 ± ·39	13·01 ± ·61
	British Neolithic ...	47·23 ± ·21 (21) [104·6, 35·3]	69·81 ± ·36 (7) [98·4, 21·3]	77·22 ± ·28 (12) [111·7, 44·1]	7·24 ± ·32 (9) [95·1, 23·6]	89·48 ± ·39	199·33 ± 1·03	122·10 ± ·43	19·15 ± ·84
Female	Spitalfields ...	3·08 ± ·18 (29) [81·1, 50·6]	2·12 ± ·29 (11) [79·8, 48·1]	4·26 ± ·24 (16) [85·3, 72·1]	1·62 ± ·26 (13) [76·1, 24·2]	4·93 ± ·28	3·54 ± ·48	5·45 ± ·31	4·41 ± ·72
	Whitechapel ...	31·82 ± ·19 (25) [80·6, 90·7]	54·96 ± ·32 (9) [78·9, 79·7]	56·55 ± ·28 (12) [84·7, 126·0]	8·99 ± ·26 (13) [76·8, 55·0]	37·28 ± ·22	69·30 ± ·40	55·81 ± ·27	14·03 ± ·41
	Farringdon Street	33·81 ± ·17 (31) [78·9, 113·9]	52·93 ± ·28 (12) [76·8, 104·7]	56·47 ± ·24 (16) [85·3, 158·1]	9·74 ± ·25 (15) [72·1, 66·7]	36·26 ± ·18	59·72 ± ·31	50·98 ± ·22	14·06 ± ·35
	Moorfields	19·46 ± ·18 (29) [81·4, 38·5]	31·45 ± ·29 (11) [80·3, 34·9]	34·66 ± ·24 (16) [85·2, 50·5]	0·76 ± ·26 (13) [76·8, 23·7]	37·24 ± ·34	66·70 ± ·59	54·65 ± ·38	2·10 ± ·73
	British Iron Age ...	9·76 ± ·22 (18) [82·1, 27·4]	17·48 ± ·39 (6) [79·8, 25·0]	16·45 ± ·30 (10) [86·3, 30·0]	1·39 ± ·34 (8) [76·7, 24·1]	23·76 ± ·55	45·91 ± 1·02	36·95 ± ·68	3·80 ± ·92
	Anglo-Saxon ...	15·31 ± ·17 (30) [78·5, 30·5]	22·53 ± ·28 (12) [76·4, 27·8]	28·04 ± ·25 (15) [85·4, 36·8]	2·57 ± ·25 (15) [71·5, 24·1]	34·88 ± ·40	55·28 ± ·67	54·51 ± ·48	7·14 ± ·68

* The number in round brackets following the coefficient is the number of characters on which it is based. The numbers in square brackets below the coefficient are the mean numbers of skulls available for the characters used in computing it: the first is for the Hythe and the second for the other series in the comparison.

but we can scarcely doubt that the best measure of racial affinity is the one which takes into account all the more important features of the skull.

No detailed comparison of single mean measurements need be made. The most significant differences between the Hythe and the British series other than the Spitalfields are found for the characters $100\,B/L$, $100\,H'/L$ and L, while B, H', U, S, Q', J and $100\,NB/NH'$ also show several high values of α. The male means of all these measurements, except H' and J, are given in Table VI of the Spitalfields paper. The Hythe series has the highest cephalic index (82·6), the English Bronze Age the second (80·9) and the Spitalfields the third highest (79·4); the height-length index is highest for the Hythe (75·4) and the English Bronze and Spitalfields have the same value (73·7), which is the next highest; the Hythe has the shortest calvarial length (177·9) and the Spitalfields the next shortest (180·7); the nasal index is highest for the Spitalfields (52·0) and next highest for the Hythe (50·7); the horizontal circumference is smallest for the Spitalfields (517·0) and next smallest for the Hythe (518·4), while the sagittal circumference is smallest for the Hythe (365·6) and next smallest for the Spitalfields (368·5). The small sizes of the calvariae of these two types distinguish them clearly from all the other British series.

(8) *Comparisons with non-British Cranial Series.* The Hythe type is seen to be closely connected with the Spitalfields and widely removed from all other British series available. It should be possible to find a number of non-British series which resemble the Hythe more closely than do all the British series except the Spitalfields. Comparisons were made with the male means of over 70 European samples and the coefficients were calculated in all cases when the connection appeared to be at all close. There are eight series in addition to the Spitalfields giving reduced coefficients for all characters less than nine, and comparisons with these are made in Table IX. Table VIII gives the mean measurements of four of these series, and all the others involved have been previously published in *Biometrika*. References to the sources from which the data relating to the Finns and Austrians (Vienna) were taken will be found in the Spitalfields paper, and the following papers provide additional series:

(i) J. Matieka: "Crânes et ossements des anciens cimetières de la ville de Prague. I. Les crânes vieux praguois du cimetière de St. Nicholas dans la vieille ville (Prague—I)." *Anthropologie (Prague)*, Vol. II, 1924, pp. 183—210 (in Czech with French résumé). The cranial series described came from a cemetery which was used from the thirteenth century until 1635. There are 115 male specimens, but facial measurements can only be given for small numbers. Means quoted in our Table VIII are for the "male" and "male?" groups combined.

(ii) J. Matieka: "On the Craniology of the Jews. I. The Skulls from the Old Cemetery (Prague—V)." *Ibid.*, Vol. IV, 1926, pp. 163—219 (in Czech with English résumé). The majority of the 53 male skulls from a single Jewish cemetery belong to the seventeenth century. The means are quoted in our Table VIII. The

type is very similar to that of the contemporary population buried in the Christian cemetery of St Nicholas.

(iii) F. Schiff: "Beiträge zur Kraniologie der Czechen." *Archiv für Anthropologie*, Bd. XXXIX (N.F. Bd. XI), 1912, S. 253—292. There are 108 male *Beinhausschädel* measured in a number of different localities in Bohemia. Means are given in *Biometrika*, Vol. XX^B, 1928, pp. 366—367, and the means of the following series are in the same place.

(iv) A. Weisbach: "Die Schädelform der Rumänen." *Denkschriften der kaiserlichen Akademie der Wissenschaften. Mathematisch-naturwissenschaftliche Classe*, Bd. XXX, 1870, S. 107—136. Measurements are given of 40 modern skulls of soldiers.

(v) C. Luigi Calori: "Del Tipo Brachicefalo negli Italiani Odierni." *Memorie dell' Accademia delle Scienze dell' Istituto di Bologna*, Serie II, Tomo VIII, 1868, pp. 205—234. Measurements are given of 100 male skulls representing the modern population of Bologna and of known sex and age, and the means calculated for these are in our Table VIII.

(vi) F. Ferraz de Macedo: *Crime et Criminel. Essai synthétique d'observations anatomiques, physiologiques, pathologiques et psychiques sur les délinquants vivants et morts selon la méthode et les procédés anthropologiques les plus rigoureux.* Lisbon, 1892. Mean measurements determined by the author are given of a collection of male skulls of 13 assassins, 25 thieves and 9 swindlers (*escrocs*) made by Lombroso at Turin and these are quoted in our Table VIII. It is probable that the majority, if not all, of these criminals were of Italian origin.

The Hythe series is seen from Table IX to have its closest connection with the Spitalfields although there are several other reduced coefficients of the same order with series from south and east-central Europe and with a series of Finns. All these relationships are much more intimate than any which can be found between the Hythe and any British series other than the Spitalfields. A similar comparison between the last and all the available European types was made, and the reduced coefficients with the Pompeians ($3{\cdot}54 \pm {\cdot}17$) and Etruscans ($4{\cdot}04 \pm {\cdot}17$) are lower than any now found. The lowest value which it is generally possible to find for these European series is of the order 3—5, so there is no reason to believe that the two English ones are peculiarly specialised. It may be noted that in the comparisons of the Spitalfields type with those most closely allied to it (see the Spitalfields paper, Table VIII) the reduced coefficients for facial characters alone were in every case greater, and in most decidedly greater, than the reduced coefficients for calvarial characters alone. Such a relation is the reverse of that generally found. From Table IX it may be seen that some of the facial coefficients with the Hythe series are less than the corresponding calvarial ones while others are greater, and the reduction is quite marked in the comparison with the Spitalfields series. The last, then, is mainly distinguished from all the continental types to which it is most closely allied by possessing some peculiar facial characters—in particular a low orbital and high nasal index—but it is mainly distinguished from the Hythe by

TABLE VIII.

*Mean Male Measurements of the Hythe and some closely allied Series**.

Character	Hythe	Italian: Bologna	Czech: Prague (Matieka)	Jewish: Prague	Italian (Criminals)
C	1456·3 (110)	1566·2 (100)†	1498·8 (45)	—	—
L	177·9 (112)	174·2 (100)	178·7 (115)	180·5 (53)	177·8 (47)
B	146·7 (112)	142·9 (100)	149·0 (115)	147·8 (52)	146·7 (47)
H'	134·1 (112)	133·6 (100)	132·5 (86)	131·2 (49)	131·8 (47)
B'	98·6 (109)	99·9 (100)	98·5 (111)	97·8 (33)	97·8 (47)
B''	124·8 (102)	—	125·3 (113)	121·4 (33)	—
Biasterionic B	112·7 (97)	—	113·2 (112)	113·3 (30)	111·5 (47)
LB	100·5 (112)	99·0 (100)	99·6 (80)	100·4 (28)	96·5 (47)
S_1'	111·4 (112)	113·3 (100)	109·7 (109)	109·1 (33)	—
S_2'	108·9 (112)	111·0 (100)	110·6 (114)	111·9 (33)	—
S_3'	96·1 (112)	93·8 (100)	91·4 (110)	94·7 (33)	—
S_1	127·2 (112)	128·5 (100)	125·4 (110)	123·9 (33)	129·1 (47)
S_2	122·3 (112)	125·7 (100)	122·9 (113)	125·4 (33)	124·1 (47)
S_3	116·2 (112)	114·1 (100)	113·7 (110)	115·6 (33)	112·3 (47)
S	365·6 (112)	368·4 (100)	367·6 (107)	365·0 (32)	365·5 (47)
U	518·4 (112)	513·8 (100)	—	—	513·3 (47)
Glabella U	—	—	518·9 (115)	520·3 (33)	—
Broca's Q'	311·3 (112)	—	320·9 (107)	308·1 (31)	316·4 (47)
fml	35·6 (111)	34·4 (100)	35·5 (83)	34·8 (24)	35·7 (47)
fmb	30·2 (110)	30·1 (100)	29·3 (88)	29·9 (24)	30·4 (47)
J	134·3 (96)	132·1 (100)	130·5 (32)	134·5 (23)	132·3 (47)
$G'H$	69·9 (89)	69·4 (100)	67·9 (33)	67·7 (24)	—
GL	94·9 (89)	—	93·9 (30)	96·2 (24)	—
GB	94·8 (99)	—	93·4 (28)	95·4 (23)	—
NH'	49·2 (99)	—	49·3 (34)	52·2 (24)	51·8 (47)
NB	24·8 (108)	—	25·0 (33)	25·3 (24)	23·6 (47)
Lacrymal O_1	39·0 (73)	38·6 (100)	37·2 (34)	40·4 (24)	38·6 (47)
O_2	33·0 (111)	32·8 (100)	32·3 (35)	33·25 (24)	33·2 (47)
SC	9·6 (109)	—	—	9·7 (23)	—
100 B/L	82·6 (112)	82·3 (100)	83·5 (115)	82·0 (53)	82·4 (47)
100 H'/L	75·4 (112)	{76·7 (100)}‡	74·4 (85)	72·6 (49)	74·0 (47)
100 B/H'	109·5 (112)	{107·0 (100)}	{112·5 (86)}	{112·7 (49)}	{111·3 (47)}
100 fmb/fml	84·9 (110)	{87·6 (100)}	83·7 (80)	86·2 (29)	85·3 (47)
$Oc.\ I.$	59·8 (112)	59·9 (100)	{57·3 (110)}	59·0 (33)	—
100 $G'H/GB$	73·5 (80)	—	71·7 (26)	71·1 (23)	—
100 NB/NH'	50·7 (95)	—	50·9 (33)	48·6 (24)	45·6 (47)
100 O_2/Lacrymal O_1	84·6 (73)	{84·9 (100)}	85·4 (33)	82·35 (24)	86·2 (47)
$N\angle$	64°·6 (89)	—	{64°·9 (30)}	{66°·3 (24)}	—
$A\angle$	73°·7 (89)	—	{74°·3 (30)}	{73°·5 (24)}	—
$B\angle$	41°·7 (89)	—	{40°·8 (30)}	{40°·2 (24)}	—

* The mean measurements of all the other series used in this paper will be found in the papers in earlier volumes of *Biometrika* cited in the Spitalfields paper.

† This capacity was not used in calculating the coefficients of racial likeness as it is almost certainly too high.

‡ Indices and angles in curled brackets were found from the means of the component lengths instead of from individual values. The occipital index found in this indirect way is generally about one unit smaller than the true mean, but there is a closer approximation for all other indices and for the angles.

TABLE IX.

Lowest Male Coefficients of Racial Likeness between the Hythe and Other Series.*

Series	Crude Coefficients				Reduced Coefficients			
	All Characters	Indices and Angles	Calvarial Characters	Facial Characters	All Characters	Indices and Angles	Calvarial Characters	Facial Characters
Spitalfields ...	5·18 ± ·17 (31) [101·3, 163·6]	7·59 ± ·28 (12) [96·3, 152·1]	8·40 ± ·24 (16) [111·4, 236·7]	1·75 ± ·25 (15) [90·5, 85·7]	4·14 ± ·14	6·43 ± ·23	5·54 ± ·16	1·99 ± ·28
Italian: Bologna	4·88 ± ·22 (18) [105·0, 100·0]	5·36 ± ·43 (5) [103·8, 100·0]	6·10 ± ·26 (13) [111·4, 100·0]	1·72 ± ·43 (5) [88·4, 100·0]	4·77 ± ·22	5·26 ± ·42	5·79 ± ·25	1·83 ± ·46
Czech (Matieka)	4·60 ± ·19 (25) [102·4, 66·2]	2·20 ± ·32 (9) [96·9, 57·6]	5·29 ± ·25 (14) [111·3, 93·1]	3·73 ± ·29 (11) [91·1, 32·1]	5·72 ± ·24	3·05 ± ·44	5·21 ± ·25	7·86 ± ·61
Czech (Schiff) ...	6·59 ± ·18 (29) [101·3, 104·2]	4·94 ± ·29 (11) [98·1, 105·5]	3·53 ± ·24 (16) [111·4, 107·5]	10·34 ± ·26 (13) [89·0, 100·2]	6·41 ± ·17	4·86 ± ·28	3·23 ± ·22	10·97 ± ·28
Rumanian ...	3·91 ± ·20 (22) [103·3, 39·9]	4·00 ± ·32 (9) [99·6, 40·0]	4·07 ± ·25 (14) [111·3, 39·9]	3·62 ± ·34 (8) [89·4, 40·0]	6·79 ± ·35	7·01 ± ·56	6·93 ± ·43	6·54 ± ·61
Jewish: Prague ...	3·48 ± ·19 (25) [102·5, 32·0]	5·61 ± ·30 (10) [98·4, 33·2]	3·62 ± ·25 (14) [111·4, 38·5]	3·29 ± ·29 (11) [91·1, 23·8]	7·13 ± ·39	11·31 ± ·61	6·33 ± ·45	8·72 ± ·76
Finn	8·49 ± ·20 (23) [105·7, 118·0]	10·49 ± ·39 (6) [103·5, 116·0]	8·54 ± ·25 (15) [111·3, 126·9]	8·41 ± ·34 (8) [95·0, 101·5]	7·61 ± ·18	9·58 ± ·36	7·20 ± ·21	8·57 ± ·34
Austrian: Vienna	5·47 ± ·21 (21) [103·0, 49·7]	7·49 ± ·32 (9) [99·6, 49·6]	5·73 ± ·26 (13) [111·4, 50·0]	5·06 ± ·34 (8) [89·5, 49·3]	8·16 ± ·31	11·32 ± ·48	8·31 ± ·38	7·96 ± ·53
Italian (Criminals)	5·42 ± ·21 (21) [105·5, 47·0]	8·12 ± ·39 (6) [102·3, 47·0]	2·80 ± ·25 (14) [111·4, 47·0]	10·65 ± ·36 (7) [93·6, 47·0]	8·33 ± ·32	12·60 ± ·60	4·23 ± ·39	17·02 ± ·58

* Footnote as for Table VII.

certain calvarial characters. The facial characters of the Hythe type are less peculiar, but they also tend to affect the coefficients for all characters more in proportion to the effect of the calvarial characters than is found for comparisons in general.

In order to make the comparison between these two closely allied English series and the continental ones as complete as possible, the coefficients of racial likeness were calculated between all possible pairs of the series which closely resemble either of the English ones. The reduced values for all characters between the Spitalfields series and the six continental series to which it is most closely allied are given in Table X of the earlier paper, and the highest value found between any pair of these seven series was 8·57. Those between the Hythe and the eight continental series to which it is most closely allied are in Table X of the present paper, and the highest value found between any pair of these nine series is 20·32, while 15 of the possible 36 comparisons show coefficients greater than 9. The Spitalfields series is fully entitled to be considered a member of a group of closely allied racial types, all the others being continental ones. The Hythe cannot be considered a member of the same group and its nearest relationships are with a number of types which are not all intimately allied to one another. The foreign series which are closely linked to both English series are those of Finns, Austrians (Vienna) and Etruscans, and in each one of these cases it is the Spitalfields which has the lower coefficient than the Hythe. A diagram illustrating the fact that the two types with which we are primarily concerned have distinctly different racial affinities, in spite of their close resemblance to one another, has been given in the Spitalfields paper (Fig. XIII). There are twelve continental series which we have selected from a much larger number on account of their affinities to the two English ones (see Table X): four of these are of Italian origin, one came from Vienna and three (including the Jewish one) from Bohemia, while another came from Rumania. These nine series represent a comparatively restricted area of Southern Europe. Two of the remaining three came from Paris and the other from Finland. The only surprising feature of this grouping is the association of the Finns with the other racial types.

We have to attempt to reconcile these results, which have been reached by using purely quantitative methods, with other evidence relating to the ethnic history of England. The Spitalfields and Hythe types are closely similar to one another and both may be considered alien in the sense that they cannot be supposed to have represented the bulk of the population of the country at any one time. It is extremely unlikely that they were intrusive in England during any prehistoric period. The fact that the Spitalfields type bears its closest resemblances to those of Pompeians and Etruscans—the connections with these being as intimate as those between two seventeenth century London series—suggests forcibly that we are dealing with the population of London in Roman times. The sample known to us was probably of pure, or almost pure, Italian origin. The Hythe crania are certainly of post-Roman date, but we suggest that they represent a population which was directly descended from the marines and auxiliaries who are known to have been stationed in the neighbourhood during the occupation. There were probably

TABLE X. *Reduced Coefficients of Racial Likeness between Male Series closely related to the Hythe and Spitalfields Series.*

Series closely related to the Hythe series	Hythe	Spitalfields	Italian: Bologna	Czech (Matieka)	Czech (Schiff)	Rumanian	Jewish: Prague	Italian (Criminals)
Hythe	—	4·14 ± ·14 (31)	4·77 ± ·22 (18)	5·72 ± ·24 (25)	6·41 ± ·17 (29)	6·79 ± ·35 (22)	7·13 ± ·39 (25)	8·33 ± ·32 (21)
Spitalfields ...	4·14 ± ·14 (31)	—	9·11 ± ·16 (19)	13·01 ± ·19 (25)	9·06 ± ·14 (29)	13·87 ± ·31 (22)	11·62 ± ·36 (24)	15·62 ± ·28 (21)
Italian: Bologna	4·77 ± ·22 (18)	9·11 ± ·16 (19)	—	14·17 ± ·26 (17)	8·66 ± ·22 (17)	10·55 ± ·42 (16)	19·81 ± ·44 (18)	7·83 ± ·36 (17)
Czech (Matieka)	5·72 ± ·24 (25)	13·01 ± ·19 (25)	14·17 ± ·26 (17)	—	4·94 ± ·24 (23)	15·40 ± ·40 (21)	10·14 ± ·42 (26)	7·92 ± ·38 (20)
Czech (Schiff) ...	6·41 ± ·17 (29)	9·06 ± ·14 (29)	8·66 ± ·22 (17)	4·94 ± ·24 (23)	—	5·10 ± ·36 (22)	7·09 ± ·41 (22)	2·84 ± ·34 (18)
Rumanian ...	6·79 ± ·35 (22)	13·87 ± ·31 (22)	10·55 ± ·42 (16)	15·40 ± ·40 (21)	5·10 ± ·36 (22)	—	16·94 ± ·58 (20)	8·93 ± ·59 (14)
Jewish: Prague	7·13 ± ·39 (25)	11·62 ± ·36 (24)	19·81 ± ·44 (18)	10·14 ± ·42 (26)	7·09 ± ·41 (22)	16·94 ± ·58 (20)	—	7·32 ± ·56 (20)
Italian (Criminals)	8·33 ± ·32 (21)	15·62 ± ·28 (21)	7·83 ± ·36 (17)	7·92 ± ·38 (20)	2·84 ± ·34 (18)	8·93 ± ·59 (14)	7·32 ± ·56 (20)	—
Series closely related to the Hythe and Spitalfields series								
Finn	7·61 ± ·18 (23)	5·47 ± ·14 (24)	12·33 ± ·21 (16)	11·83 ± ·22 (21)	11·79 ± ·18 (24)	10·82 ± ·36 (19)	9·61 ± ·40 (20)	11·72 ± ·31 (22)
Austrian: Vienna	8·16 ± ·31 (21)	7·41 ± ·26 (22)	18·47 ± ·36 (16)	6·67 ± ·35 (21)	11·68 ± ·30 (22)	20·32 ± ·42 (27)	4·79 ± ·50 (20)	2·60 ± ·52 (14)
Series closely related to the Spitalfields series								
Pompeian ...	12·34 ± ·21 (23)	3·54 ± ·17 (23)	15·50 ± ·26 (18)	14·85 ± ·27 (22)	9·97 ± ·21 (23)	17·58 ± ·40 (19)	11·60 ± ·43 (22)	6·20 ± ·35 (20)
Etruscan ...	8·45 ± ·20 (29)	4·04 ± ·17 (29)	21·69 ± ·25 (18)	17·52 ± ·27 (24)	14·52 ± ·22 (25)	10·07 ± ·41 (19)	13·90 ± ·43 (23)	11·55 ± ·36 (20)
Parisian: Cité ...	13·52 ± ·30 (22)	4·28 ± ·24 (22)	17·89 ± ·31 (18)	20·55 ± ·33 (21)	11·89 ± ·29 (19)	21·67 ± ·51 (15)	8·49 ± ·50 (21)	14·15 ± ·38 (24)
Parisian: L'Ouest	11·38 ± ·24 (20)	8·57 ± ·19 (20)	19·32 ± ·27 (16)	16·38 ± ·29 (19)	15·25 ± ·26 (17)	26·04 ± ·50 (13)	6·50 ± ·47 (19)	7·84 ± ·37 (20)

merchants settled there at this time as well. Judging from the affinities of their presumed descendants, these people were partly of Italian and partly of East European origin, the closest affinities of the later population being found with modern types from Bologna, Bohemia and Rumania. The inhabitants of Hythe, and of the castle and port at Lympne, may have formed a racially heterogeneous populace in Roman times, but the variability of later generations would have been reduced by intermarriage. It is probable that the Italian population of London which remained was exterminated, or absorbed by larger numbers of new-comers, in the Saxon period, but it appears to us not improbable that the Roman population of Hythe persisted, with no appreciable modification from outside, until late mediaeval times at least. This peculiar racial type was less likely to remain pure as the town declined in size, but it may only have been completely absorbed in recent times by that one to which the vast majority of the present-day inhabitants of England conform. These appear to us to be the most reasonable hypotheses which will reconcile all the evidence at present available, but a more extended anthropometric survey of the past and present populations of England would be needed in order to make theories of this kind at all definite.

(9) *Conclusions.* The age of the human skeletons preserved in the ambulatory passage of St Leonard's Church, Hythe, is not known exactly. The earliest direct reference to them is found in a book written in 1678, and they appear to have been shown regularly to visitors from some years before that until the present day. The church is an early Norman one and it was enlarged considerably in Late Norman times and again at the beginning of the thirteenth century. The passage was built under the chancel at this last date with the object, it is thought, of preserving a processional way on consecrated ground all round the outside of the building. It is well lighted and above ground, and these circumstances account for the fact that it has remained open to inspection. Various theories associating the skeletons with battles or massacres have been broached at different times, but all these are quite untenable, and two of the battles referred to were probably not fought anywhere near Hythe. Men and women are represented approximately in the proportion 1·3 to 1 and there are also a few juvenile individuals present. There are said to be more than 8000 femora in the collection, and there are probably more than 2000 skulls, of which half are nearly complete. Other bones of the skeleton, however, can only be found in small numbers, and it is clear that there was a stringent selection in favour of the crania and femora when the collection was made. Everything points to the fact that these are graveyard bones and no other plausible explanation of their origin can be offered. The town was at one time considerably larger than it is to-day. There were at least three churches within the present-day limits of Hythe and another about a mile away. All these, except St Leonard's, had been abandoned by the end of the fourteenth century. The disused graveyards do not appear to have been built on since, but bones may have been disinterred from them at various times. It is known that three skulls were taken from one of the sites and placed in the ambulatory as recently as 1912. The majority of the remains were

probably dug up from the ground of St Leonard's and it is quite likely that the nucleus of the collection was placed in the ambulatory shortly after it had been built, since this extension of the church almost certainly covered the original graveyard. When the practice had once been started it may well have continued for some centuries, but the vast majority, if not all, of the bones must have been collected before 1650. It is most probable that the chamber was used both as a charnel-house and as an ambulatory at the same time. The suggestion that it would not have been used for the former purpose until the processions were discontinued at the Reformation has been made. This would leave only the hundred years from 1550 to 1650 for the bulk of the material to have been collected. The yearly burials in this period are known to have been between 30 and 40 and the normal disinterments must have been supplemented by the addition of at least 1000 skeletons dug up at one particular time in order to give the total now present. It appears to us more probable that the collection was made between 1250 and 1650 and that the people represented died between 1100 and 1600. These dates are, of course, only approximate and it is just possible that an examination of the not yet edited records of Hythe might render them more precise.

The *Portus Lemanis*, one of the principal ports of the Saxon Shore, was at West Hythe, two miles from St Leonard's Church, and it was protected by the large castrum which became known later as Stutfall Castle. Roman remains have also been found in Hythe itself, and there must have been a considerable population in the district during the occupation. Little is known about the town in the Saxon period, but it became a Cinque Port about the time of the Norman Conquest. The primary cause of its decline was the silting up of the harbour which once extended from West Hythe to Sandgate. This gradual process had made the coast entirely unsuited for a large port before 1600. Fire, plague and the ravages of French seamen hastened the decline of the population.

The first anthropological description of the Hythe crania which is of any value is that of Professor F. G. Parsons published in 1908. He gives a few measurements of 590 specimens. We have dealt with a separate sample of 199 crania, and all the usual measurements, type contours and remarks are provided for these. The two independent series might be supposed to have been drawn at random from the original material, except that preference was given to the more complete specimens in our case. The mean measurements of the female series show no differences of any consequence, but Professor Parsons found a male calvarial breadth which is significantly less than ours, and the male cephalic indices are as a result also differentiated. It is shown that these differences are not due to instrumental errors, or to differences of technique. Hypotheses are suggested to account for this deviation in a single character between the crania shelved in the south bay (Parsons) and in the north bay (Stoessiger and Morant) of the ambulatory *.

Our series shows as small a variability as any other European series. Comparisons between it and other European series were made by Professor Karl Pearson's

* See Note, p. 202.

method of the coefficient of racial likeness. The Hythe type was found to be widely removed from all other British types, except that of the crania recently excavated at Spitalfields Market, London, and these two resemble one another closely. Their cephalic indices (82·6 and 79·4 respectively for male skulls) are both distinctly higher than all other English mean values except those of the Bronze Age (80·9), but the last skull is of a larger type than the other two and cannot be supposed to be closely allied to them. The Hythe is more nearly related to the Spitalfields than to any other series with which it can be compared, and the connection is as close as the most intimate which are normally met with among European series. Relationships of the same order are found with modern types representing the populations of Bologna, Bohemia and Rumania. The Spitalfields series has rather closer relationships with both Pompeians and Etruscans than with the mediaeval inhabitants of Hythe. There appears to be only one plausible explanation of these facts. Both the types must be supposed alien to England, in so far as they do not represent the bulk of the population of this country at any one period. The Spitalfields interments are very possibly of Roman date and they are probably those of people of pure, or nearly pure Italian origin. The Hythe skeletons are certainly of post-Roman date, but they are probably those of people who were directly descended from the foreigners who lived in the district during the occupation. These foreigners may be supposed to have come partly from Italy and partly from Central Europe or the Balkan area, but their descendants would show less variability owing to inter-marriage. There is good reason to believe that the Roman element in the population of London had been eliminated before mediaeval times, but this was not so at Hythe. Measurements of the present-day inhabitants suggest, however, that the type which lingered on for several centuries has at last become transformed into one which appears to be spread uniformly over the country at present. More extended anthropometric investigations of the existing and past populations of England will be needed in order to test the validity of these hypotheses.

DESCRIPTION OF PLATES.

I. Engravings of the Interior of the Ambulatory Passage of St Leonard's Church, Hythe.

(*a*) By Thomas Russell, 1783. The legend associates the bones with the battle between the Britons and Saxons in 456 A.D. The stack of bones is said to be thirty feet long, eight feet high and eight feet "over." (*b*) By W. Deeble. From: *Delineations, Historical and Topographical of the Isle of Thanet and the Cinque Ports*, Vol. II. 1818. By E. W. Brayley. (*c*) By G. Rowe, N.D. The drawing was probably made shortly after the skulls were first arranged in 1851 on the shelves.

II. Typical Hythe Skulls. *Norma lateralis.*

(*a*) No. 1043. Male. (*b*) No. 758. Female.

III. Typical Hythe Skulls. *Norma facialis.*

(*a*) No. 1043. Male. (*b*) No. 758. Female.

IV. Typical Hythe Skulls. *Norma verticalis.*

(*a*) No. 1043. Male. (*b*) No. 758. Female.

V. **Abnormal Hythe Skulls.**

(a) No. 929. Male. Eminence at vertex. (b) No. 967. Female. Unusual globular form, possibly hydrocephalous.

VI. **Basal Anomalies of the Hythe Skulls.**

(a) No. 1004. Female. Fossa pharyngea and irregular anterior border of foramen. (b) No. 659. Female. Opisthial notch. (c) No. 1071. Male. Thickened right side of basi-occipital. (d) No. 771. Female. Articular surface at basion.

VII. **Anomalous Hythe Skulls.**

(a) No. 620. Female. Canine erupting behind incisors. (b) No. 646. Male. Anomalous dentition on right side. (c) No. 867. Female. Ex-occipital separate on left side. (d) No. 895. Male. Arthritic glenoid surface and large tympanic perforation. (e) and (f) No. 1054. Female. Anomalous auricular passages with irregular posterior walls and multiple perforations of the tympanic plates.

Note (May 20th, 1932). The fact that significant differences are found between the samples of the Hythe crania measured by Professor Parsons and by the present writers was of sufficient importance to make further inquiry desirable. He measured the total 556 skulls on the shelves in the south bay of the ambulatory, and on re-measuring 117 of these we found a close agreement between his and our individual readings. The 199 more complete crania on the shelves in the north bay were selected from the total of about 500 there, and the selected group is described in detail in the present paper. Table I above shows that there are significant differences between the two principal samples from the north and south bays respectively in the case of the mean male breadths and cephalic indices. These might have been due to the fact that our sample was selected while that dealt with by Professor Parsons was not. To decide this point we have recently measured the lengths and breadths of about three-quarters of the remaining (and hence less complete) crania on the shelves in the north bay for which the cephalic index can be found. The means for this sample are:

	L	B	$100B/L$
♂	179·1 ± ·40 (110)	146·0 ± ·38 (110)	81·6 ± ·27 (110)
♀	171·3 ± ·44 (76)	141·3 ± ·44 (76)	82·6 ± ·31 (76)

Thus the means of the two samples from the north bay agree closely with one another, and it must be concluded that the crania in the south bay have a slightly lesser mean breadth and cephalic index than those of the crania in the north bay. We infer that the original pile contained differentiated strata, whether due to a secular change in the Hythe population, or to the influx of another race it is not possible to determine.

(*a*) 1783.

(*b*) 1818.

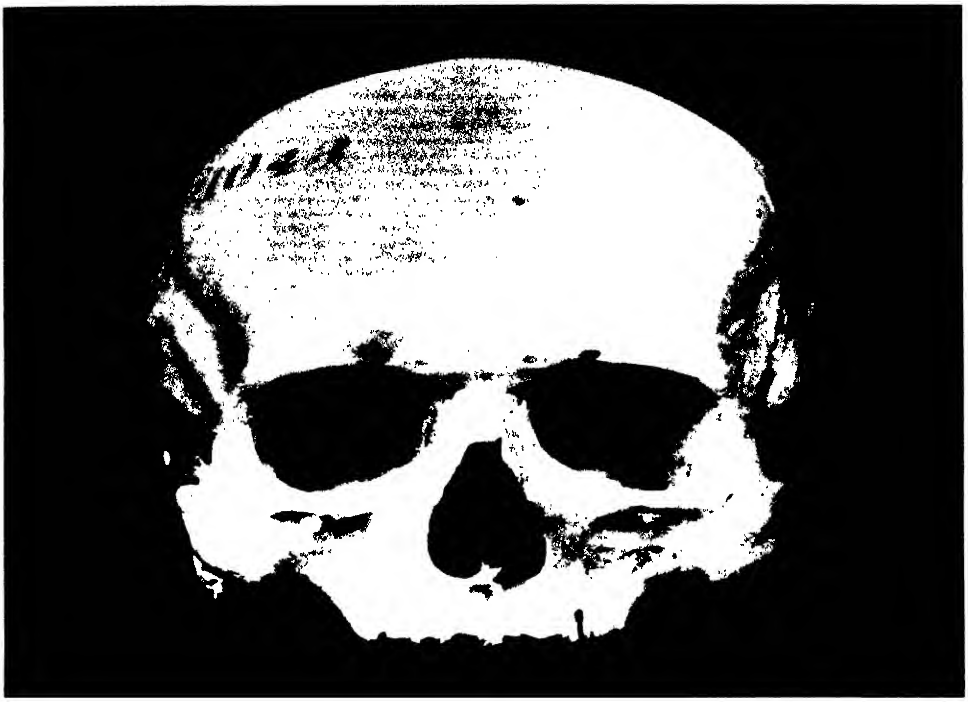
(*c*) After 1851.

Engravings of the Interior of the Ambulatory Passage of St Leonard's Church, Hythe

(*a*) No. 1013. Male.

(*b*) No. 758. Female.

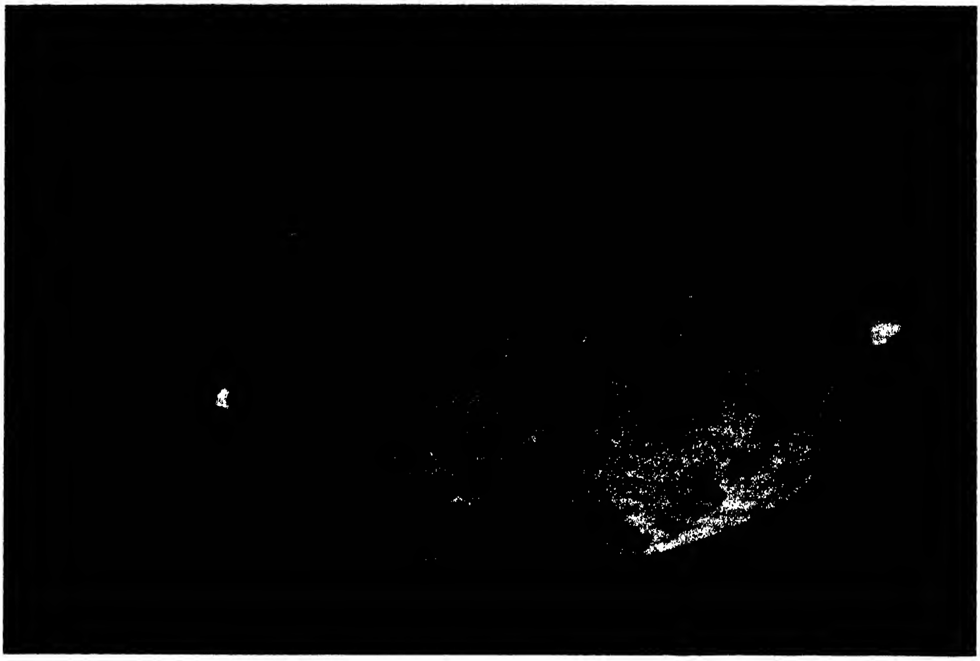
Typical Hythe Skulls. *Norma lateralis* (ca. 0·6 natural size)



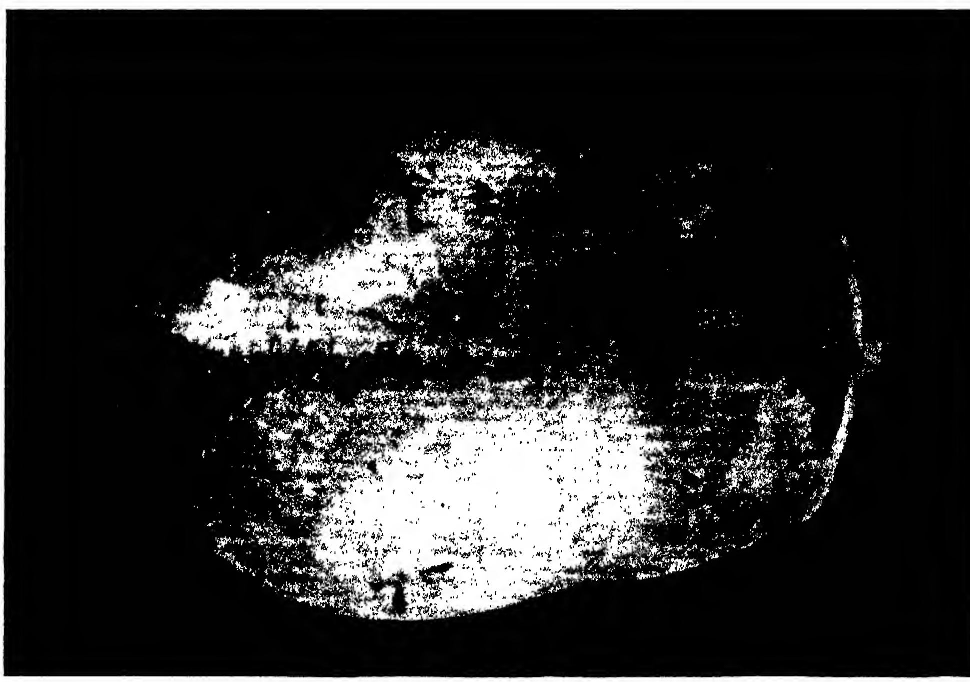
(*a*) No. 1043. Male.

(*b*) No. 758. Female.

Typical Hythe Skulls. *Norma facialis* (ca. 0·6 natural size)



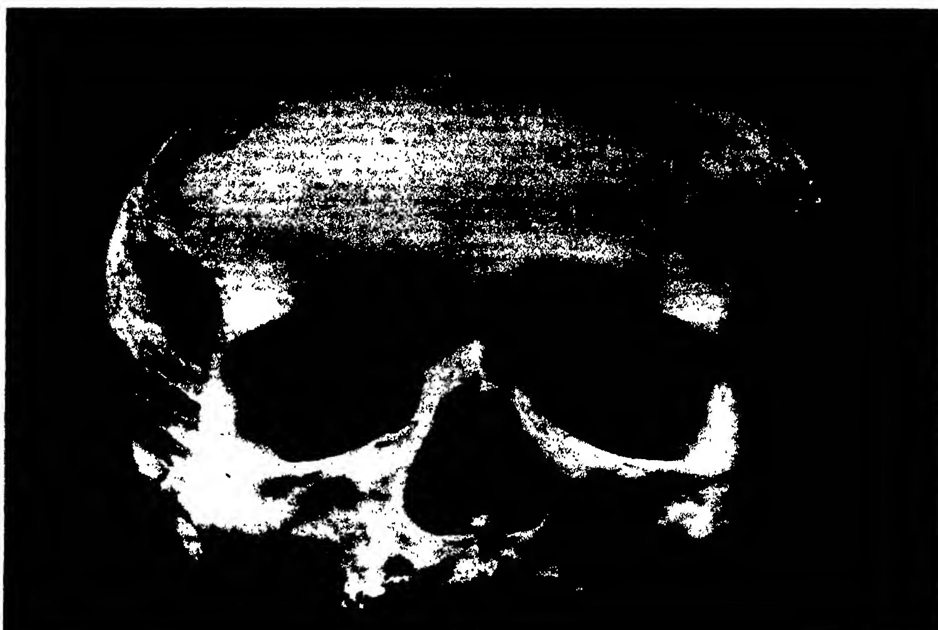
(*a*) No. 1043. Male (ca. 0·6 natural size).



(*b*) No. 758. Female (ca. 0·65 natural size).

Typical Hythe Skulls. *Norma verticalis*

(*a*) No. 929. Male. Eminence at vertex.



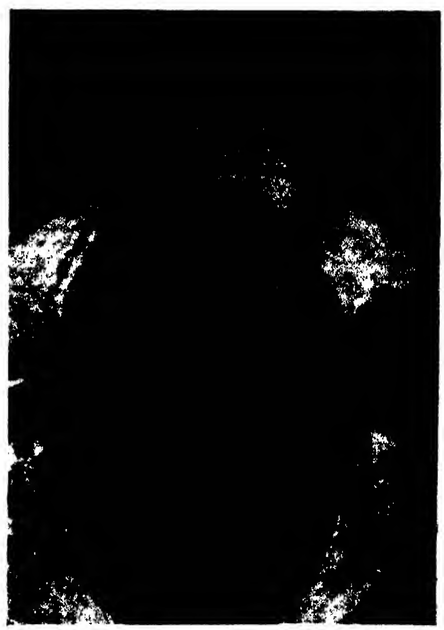
(*b*) No. 967. Female. Unusual globular form, possibly hydrocephalous.

Abnormal Hythe Skulls (ca. 0·6 natural size)

(*a*) No. 1004. Female. Fossa pharyngea and irregular anterior border of foramen.

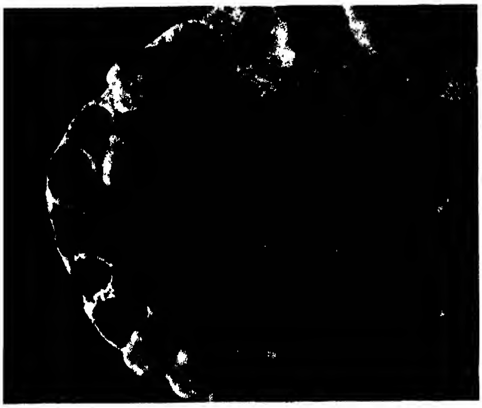
(*b*) No. 659. Female. Opisthial notch.

(*c*) No. 1071. Male. Thickened right side of basi-occipital.



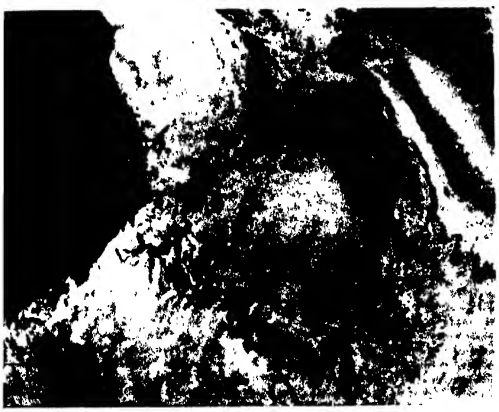
(*d*) No. 771. Female. Articular surface at basion.

Basal Anomalies of the Hythe Skulls (ca. 1·4 natural size)



(*a*) No. 620, ca. 1·3 natural size. Female. Canine erupting behind incisors.

(*b*) No. 646, ca. 1·2 natural size. Male. Anomalous dentition on right side.



(*c*) No. 867, ca. 1·5 natural size. Female. Ex-occipital separate on left side.

(*d*) No. 895, ca. 1·5 natural size. Male. Arthritic glenoid surface and large tympanic perforation.

(*e*) and (*f*) No. 1054, ca. 1·9 natural size. Female. Anomalous auricular passages with irregular posterior walls and multiple perforations of the tympanic plates.

Anomalous Hythe Skulls

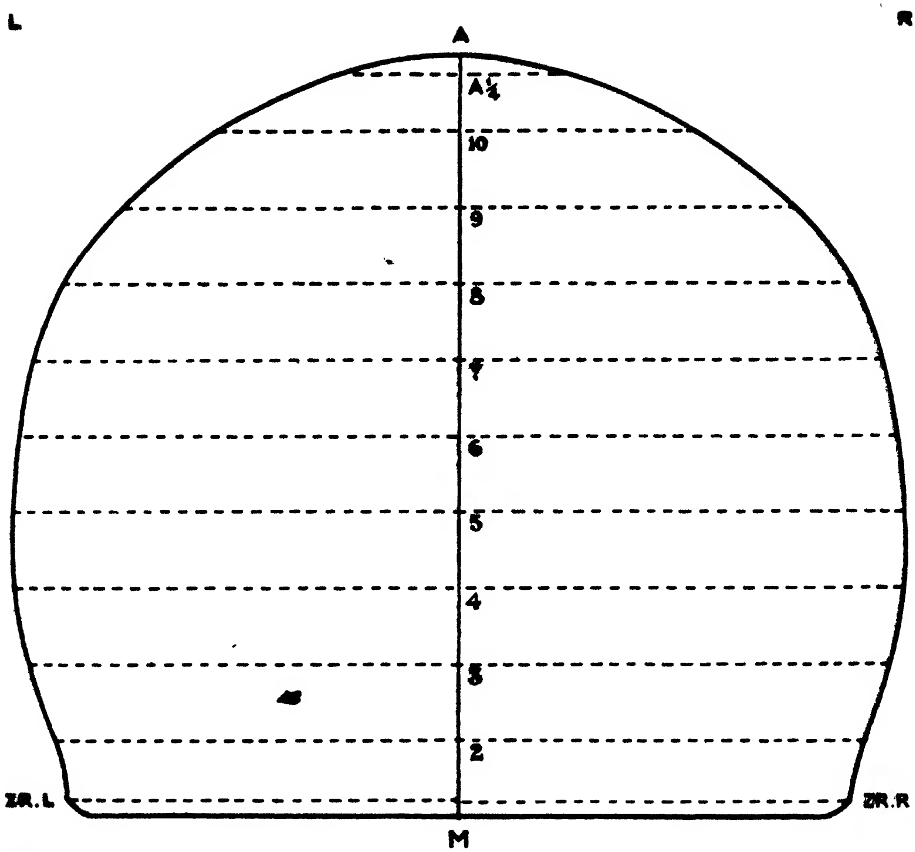


Fig. III Transverse Type Contour of 112 ♂ Hythe Skulls.

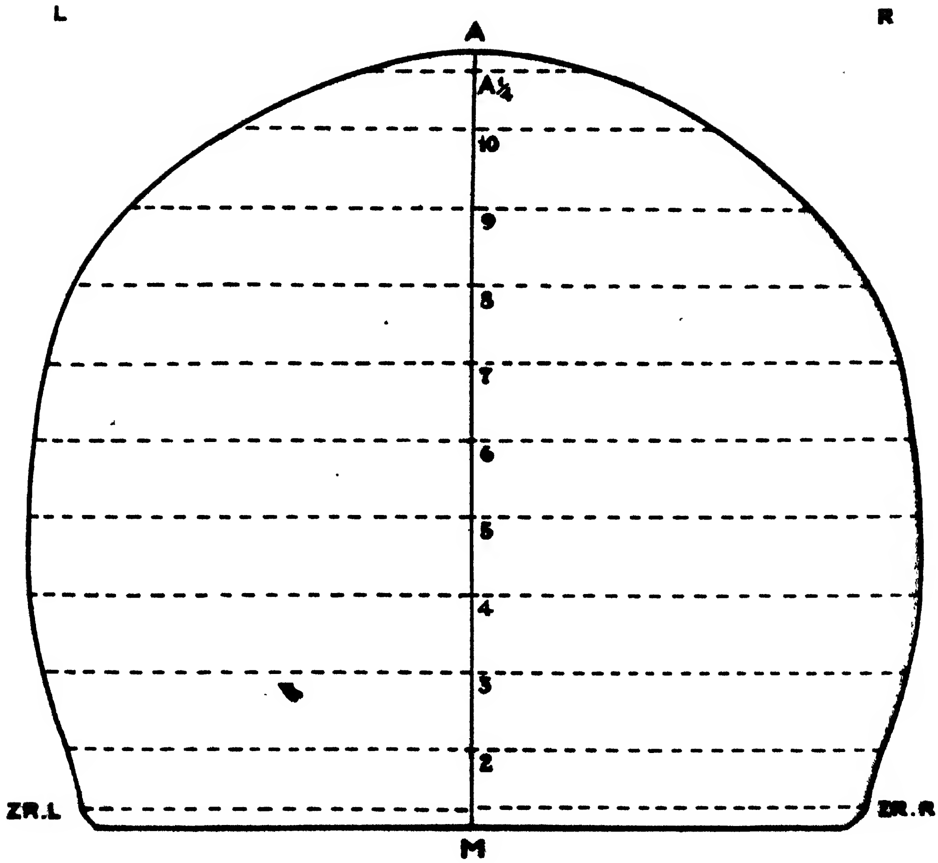


Fig. III Transverse Type Contour of 86 ♀ Hythe Skulls.

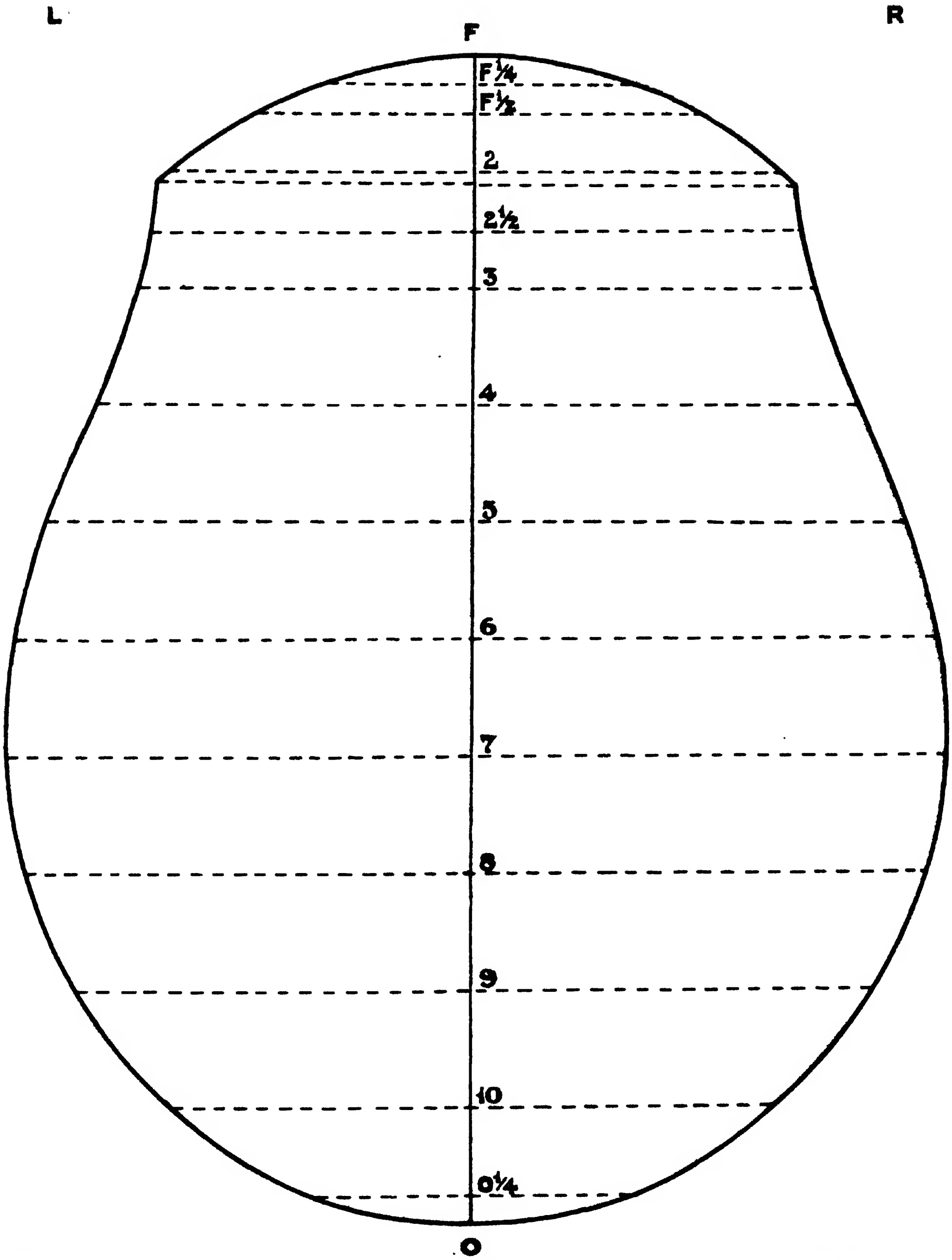
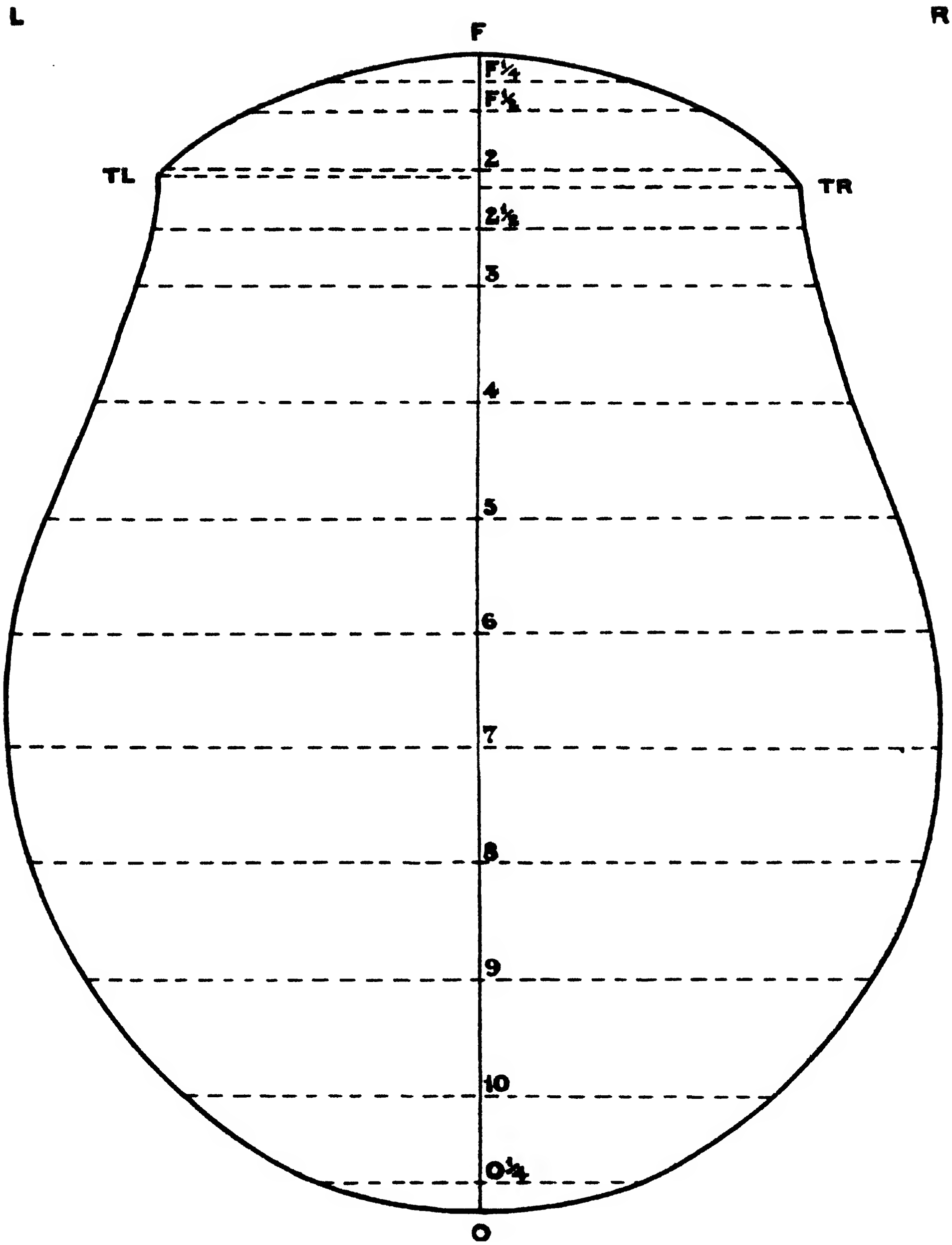


Fig. V Horizontal Type Contour of 112 ♂ Hythe Skulls



: Horizontal Type Contour of 86 ♀ Hythe Skulls.

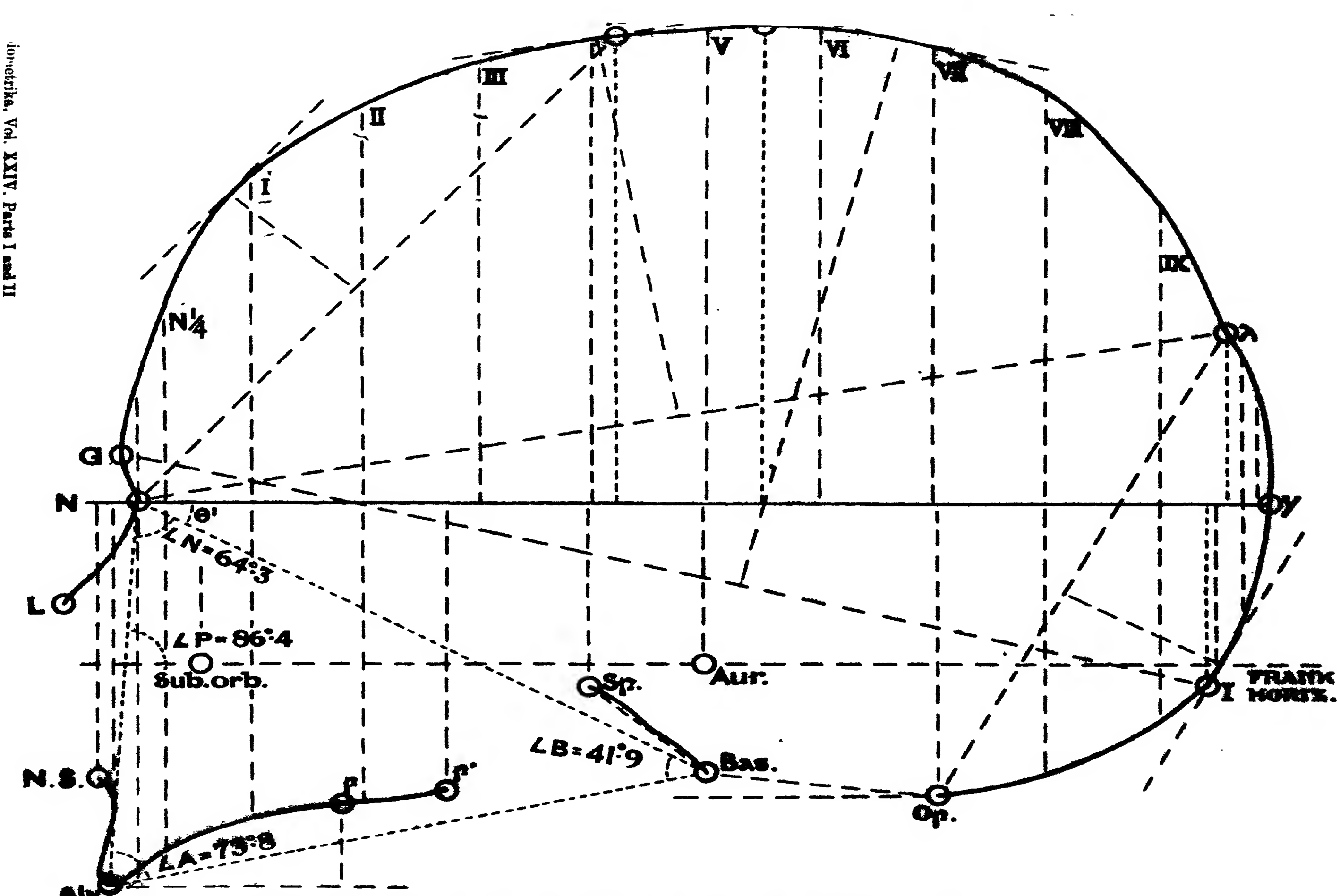


Fig. VII Sagittal Type Contour of 112 ♂ Hythe Skulls.

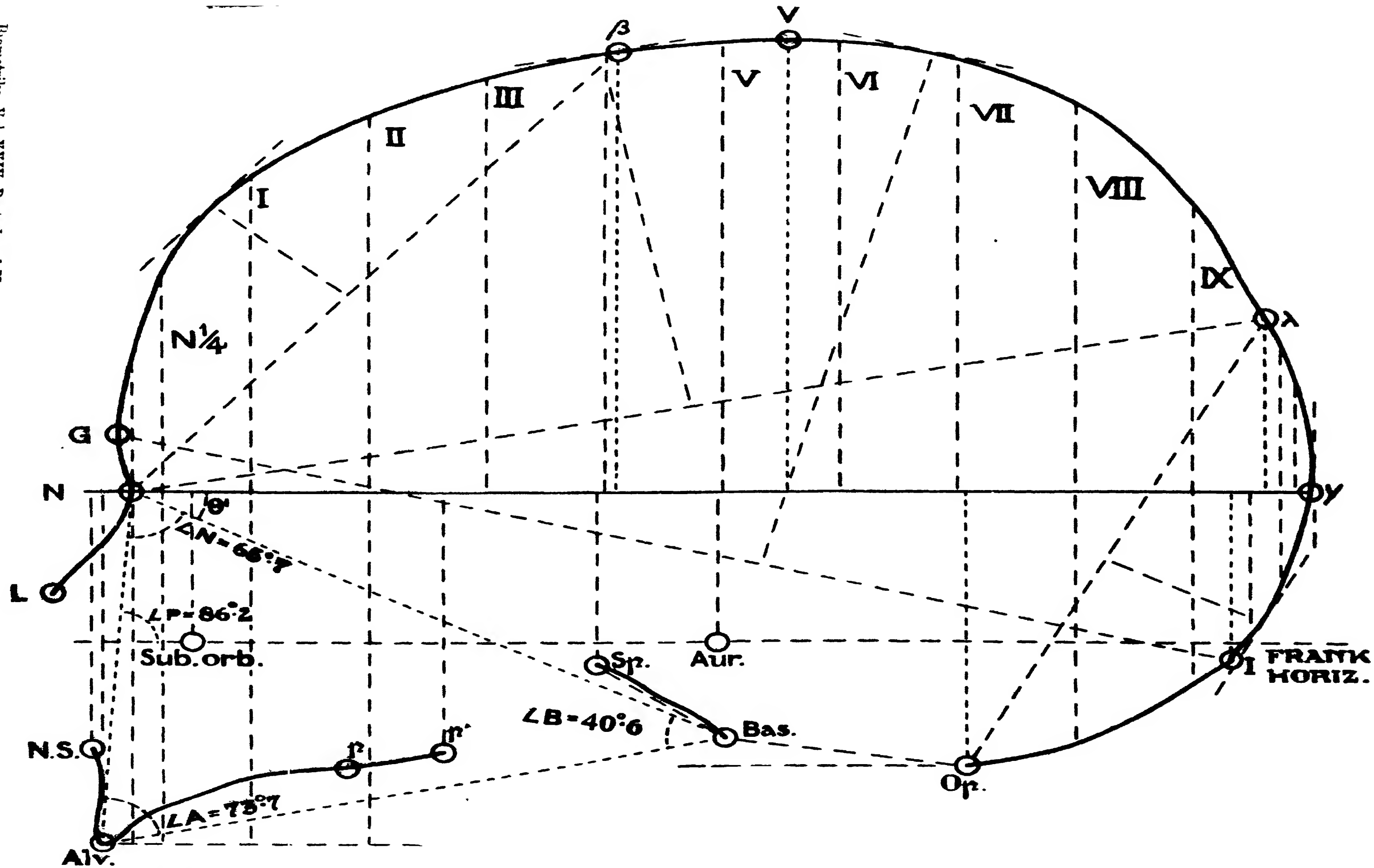


Fig. VIII Sagittal Type Contour of 87 ♀ Hythe Skulls.

ON THE MEAN CHARACTER AND VARIANCE OF A RANKED INDIVIDUAL, AND ON THE MEAN AND VARIANCE OF THE INTERVALS BETWEEN RANKED INDIVIDUALS.

PART II. CASE OF CERTAIN SKEW CURVES.

BY KARL PEARSON WITH THE ASSISTANCE OF MARGARET V. PEARSON.

(1) IN an earlier paper* I dealt at length with the case of Ranks and Rank-Intervals in sampling from a normal distribution. The methods were approximate, but sufficed to link up the accurate expressions of Professor Hojo† for small samples with the asymptotic values for large samples given by me in 1920‡.

In the present paper I propose to consider the influence of skewness in the parent distributions on the distribution of ranks, but in doing so confine myself at present to certain special skew-curves, which lie, however, rather widely distributed over the (β_1, β_2) plane, and have accordingly a very considerable range of skewness. The selection of these curves for the second part of this paper has been made, in the first place, because the ranking problem admits in their cases of exact solution. I shall start with the exponential curve $y = \frac{N}{\sigma} e^{-\frac{x}{\sigma}}$, where the mean is at a distance σ, the standard deviation, from the origin§. I shall state in the first place certain preliminary propositions relating to the B-function.

(2) *On certain Functions allied to the* B*-function.*

The well-known relation $\quad \mathrm{B}(q, q') = \dfrac{\Gamma(q)\,\Gamma(q')}{\Gamma(q+q')}$

leads us at once to

$$\frac{d \log \mathrm{B}(q, q')}{dq'} = \frac{d \log \Gamma(q')}{dq'} - \frac{d \log \Gamma(q+q')}{dq'},$$

and generally

$$\frac{d^t \log \mathrm{B}(q, q')}{dq'^t} = \frac{d^t \log \Gamma(q')}{dq'^t} - \frac{d^t \log \Gamma(q+q')}{dq'^t}.$$

Now

$$\frac{d \log \Gamma(1+x)}{dx} = -\xi + \tfrac{1}{1} + \tfrac{1}{2} + \tfrac{1}{3} + \dots + \frac{1}{x},$$

where ξ = Euler's constant = ·5772,1566,4901,53.

* *Biometrika*, Vol. XXIII. pp. 364—397.

† *Ibid.* Vol. XXIII. pp. 315—360.

‡ *Ibid.* Vol. XIII. pp. 113—32.

§ We have $\beta_1 = 4$, $\beta_2 = 9$.

Let us consider the special case of

$$\mathrm{B}(q, n-q+1) = \frac{\Gamma(q)\,\Gamma(n-q+1)}{\Gamma(n+1)};$$

then

$$\frac{d \log \mathrm{B}(q, n-q+1)}{dn} = \frac{d \log \Gamma(n-q+1)}{d(n-q)} - \frac{d \log \Gamma(n+1)}{dn}$$

$$= -\left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \dots + \frac{1}{n-q+1}\right)$$

$$= -\underset{s=0}{\overset{q-1}{S}} \frac{1}{n-s},$$

$$\frac{d^2 \log \mathrm{B}(q, n-q+1)}{dn^2} = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \frac{1}{(n-2)^2} + \dots + \frac{1}{(n-q+1)^2}$$

$$= \underset{s=0}{\overset{q-1}{S}} \frac{1}{(n-s)^2},$$

$$\frac{d^3 \log \mathrm{B}(q, n-q+1)}{dn^3} = -2 \underset{s=0}{\overset{q-1}{S}} \frac{1}{(n-s)^3},$$

and generally

$$\frac{d^t \log \mathrm{B}(q, n-q+1)}{dn^t} = (-1)^t\,\Gamma(t) \underset{s=0}{\overset{q-1}{S}} \frac{1}{(n-s)^t}.$$

I propose to term the functions of B in the above series the dibeta, tribeta, tetrabeta, pentabeta, etc. functions, corresponding to the similar functions of the Γ-function, to which in 1919 I gave the names of digamma, trigamma, etc. functions. The relations between these log beta and the log gamma functions are very simple, but it is well to have separate names for them, and symbols likewise, as this much simplifies the work we have in hand.

As the capital beta is not a convenient letter to modify I adopt the older Greek form of β, that is $\mathfrak{B}$, which is in itself of interest as showing the evolution of that letter in its capital form. I wish it had originally been adopted to represent the B-function itself. Adopting the symbols of 1919 we have

Γ	Gamma Function,	$\mathfrak{B}$	Beta Function,
$\digamma$	Digamma Function,	$\mathfrak{B}_1$	Dibeta Function,
$\digamma_2$	Trigamma Function,	$\mathfrak{B}_2$	Tribeta Function,
$\digamma_3$	Tetragamma Function,	$\mathfrak{B}_3$	Tetrabeta Function,
$\digamma_4$	Pentagamma Function,	$\mathfrak{B}_4$	Pentabeta Function.

The relations between these functions are exceedingly simple. We have

$$\mathfrak{B}(q, n-q+1) = \Gamma(q)\,\Gamma(n-q+1)/\Gamma(n+1),$$

$$\mathfrak{B}_1(q, n-q+1) = \digamma(n-q+1) - \digamma(n+1),$$

$$\mathfrak{B}_2(q, n-q+1) = \digamma_2(n-q+1) - \digamma_2(n+1),$$

$$\mathfrak{B}_3(q, n-q+1) = \digamma_3(n-q+1) - \digamma_3(n+1),$$

$$\mathfrak{B}_4(q, n-q+1) = \digamma_4(n-q+1) - \digamma_4(n+1),$$

and so on.

Thus the various $\log \mathfrak{B}(q, n-q+1)$ differential coefficients can always be obtained from the corresponding $\log \Gamma$ differential coefficients, and a table of the latter will suffice to determine the former.

For our present purposes we need these functions for integer values only of n and q, and tables have been computed of inverse powers and sums of inverse powers for the purposes of this paper. They give twelve figure values from $n=1$ to 100. If the sum of the tth powers of the inverse numbers be denoted by $S\left(\frac{1}{n^t}\right)$ from n equal 1 to n, then

$$\begin{aligned}
\text{Dibeta} &= \mathfrak{B}(q, n-q+1) = -\left(\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-q+1}\right) \\
&= -\left(S\left(\frac{1}{n}\right) - S\left(\frac{1}{n-q}\right)\right), \\
\text{Tribeta} &= \mathfrak{B}(q, n-q+1) = S\left(\frac{1}{n^2}\right) - S\left(\frac{1}{n-q}\right)^2, \\
\text{Tetrabeta} &= \mathfrak{B}(q, n-q+1) = -2\left(S\left(\frac{1}{n^3}\right) - S\left(\frac{1}{(n-q)^3}\right)\right), \\
\text{Pentabeta} &= \mathfrak{B}(q, n-q+1) = 6\left(S\left(\frac{1}{n^4}\right) - S\left(\frac{1}{(n-q)^4}\right)\right),
\end{aligned}$$

and so on.

The following tables give the $S\left(\frac{1}{n^t}\right)$ values. It is scarcely necessary to remark that the $\frac{1}{n^t}$ column gives the first differences, but they are not placed here for interpolation purposes as we are concerned only with integer values of n and q. But as they have been calculated to obtain $S\left(\frac{1}{n^t}\right)$, it seemed worth while putting them on record, as the reciprocals of powers of integers are often valuable for other purposes to the computer.

For values of n outside the limits of our table, we have with ample accuracy:

$$S\left(\frac{1}{n}\right) = \cdot 5772{,}1566{,}4901{,}53 + 2\cdot 3025{,}8509{,}2994{,}05 \log_{10} n + \tfrac{1}{2}n^{-1} - \tfrac{1}{12}n^{-2} + \tfrac{1}{120}n^{-4} - \tfrac{1}{210}n^{-6},$$

$$S\left(\frac{1}{n^2}\right) = 1\cdot 6449{,}3406{,}6848{,}23 - n^{-1} + \tfrac{1}{2}n^{-2} - \tfrac{1}{6}n^{-3} + \tfrac{1}{30}n^{-5} - \tfrac{1}{42}n^{-7} + \tfrac{1}{30}n^{-9},$$

$$S\left(\frac{1}{n^3}\right) = 1\cdot 2020{,}5690{,}3159{,}59 - \tfrac{1}{2}n^{-2} + \tfrac{1}{2}n^{-3} - \tfrac{1}{4}n^{-4} + \tfrac{1}{12}n^{-6} - \tfrac{1}{12}n^{-8} + \tfrac{3}{20}n^{-10} - \tfrac{5}{12}n^{-12},$$

$$S\left(\frac{1}{n^4}\right) = 1\cdot 0823{,}2323{,}3711{,}14 - \tfrac{1}{3}n^{-3} + \tfrac{1}{2}n^{-4} - \tfrac{1}{3}n^{-5} + \tfrac{1}{6}n^{-7} - \tfrac{2}{9}n^{-9} + \tfrac{1}{2}n^{-11} - \tfrac{5}{3}n^{-13}.$$

For our analysis of ranking from the exponential curve we require to express the differential coefficients of the $\mathfrak{B}$-function in terms of the differential coefficients of the $\log \mathfrak{B}$-function. These are given on p. 215, Equations (xviii)—(xxi).

TABLE I.

Table for finding the Derived Beta and Gamma Functions.

n	Δ $1/n$	Dibeta $S(1/n)$	Δ $1/n^2$	Tribeta $S(1/n^2)$
1	1·0000 0000 0000	1·0000 0000 0000	1·0000 0000 0000	1·0000 0000 0000
2	·5000 0000 0000	1·5000 0000 0000	·2500 0000 0000	1·2500 0000 0000
3	·3333 3333 3333	1·8333 3333 3333	·1111 1111 1111	1·3611 1111 1111
4	·2500 0000 0000	2·0833 3333 3333	·0625 0000 0000	1·4236 1111 1111
5	·2000 0000 0000	2·2833 3333 3333	·0400 0000 0000	1·4636 1111 1111
6	·1666 6666 6667	2·4500 0000 0000	·0277 7777 7778	1·4913 8888 8889
7	·1428 5714 2857	2·5928 5714 2857	·0204 0816 3265+	1·5117 9705 2154
8	·1250 0000 0000	2·7178 5714 2857	·0156 2500 0000	1·5274 2205 2154
9	·1111 1111 1111	2·8289 6825 3968	·0123 4567 9012	1·5397 6773 1167
10	·1000 0000 0000	2·9289 6825 3968	·0100 0000 0000	1·5497 6773 1167
11	·0909 0909 0909	3·0198 7734 4877	·0082 6446 2810−	1·5580 3219 3976
12	·0833 3333 3333	3·1032 1067 8211	·0069 4444 4444	1·5649 7663 8421
13	·0769 2307 6923	3·1801 3375 5134	·0059 1715 9763	1·5708 9379 8184
14	·0714 2857 1429	3·2515 6232 6562	·0051 0204 0816	1·5759 9583 9001
15	·0666 6666 6667	3·3182 2899 3229	·0044 4444 4444	1·5804 4028 3445−
16	·0625 0000 0000	3·3807 2899 3229	·0039 0625 0000	1·5843 4653 3445−
17	·0588 2352 9412	3·4395 5252 2641	·0034 6020 7612	1·5878 0674 1057
18	·0555 5555 5556	3·4951 0807 8196	·0030 8641 9753	1·5908 9316 0811
19	·0526 3157 8947	3·5477 3965 7144	·0027 7008 3102	1·5936 6324 3913
20	·0500 0000 0000	3·5977 3965 7144	·0025 0000 0000	1·5961 6324 3913
21	·0476 1904 7619	3·6453 5870 4763	·0022 6757 3696	1·5984 3081 [illegible]
22	·0454 5454 5455−	3·6908 1325 0217	·0020 6611 5702	1·6004 969[illegible]
23	·0434 7826 0870	3·7342 9151 1087	·0018 9035 9168	1·6023 8729 [illegible]
24	·0416 6666 6667	3·7759 5817 7754	·0017 3611 1111	1·6041 2340 3591
25	·0400 0000 0000	3·8159 5817 7754	·0016 0000 0000	1·6057 2340 3591
26	·0384 6153 8462	3·8544 1971 6215+	·0014 7928 9941	1·6072 0269 3532
27	·0370 3703 7037	3·8914 5675 3252	·0013 7174 2112	1·6085 7443 5644
28	·0357 1428 5714	3·9271 7103 8966	·0012 7551 0204	1·6098 4994 5848
29	·0344 8275 8621	3·9616 5379 7587	·0011 8906 0642	1·6110 3900 6490
30	·0333 3333 3333	3·9949 8713 0920	·0011 1111 1111	1·6121 5011 7602
31	·0322 5806 4516	4·0272 4519 5437	·0010 4058 2726	1·6131 9070 0328
32	·0312 5000 0000	4·0584 9519 5437	·0009 7656 2500	1·6141 6726 2828
33	·0303 0303 0303	4·0887 9822 5740	·0009 1827 3646	1·6150 8553 6473
34	·0294 1176 4706	4·1182 0999 0445+	·0008 6505 1903	1·6159 5058 8377
35	·0285 7142 8571	4·1467 8141 9017	·0008 1632 6531	1·6167 6691 4907
36	·0277 7777 7778	4·1745 5919 6795−	·0007 7160 4938	1·6175 3851 9845+
37	·0270 2702 7027	4·2015 8622 3822	·0007 3046 0190	1·6182·6898 0035+
38	·0263 1578 9474	4·2279 0201 3295+	·0006 9252 0776	1·6189 6150 0811
39	·0256 4102 5641	4·2535 4303 8936	·0006 5746 2196	1·6196 1896 3007
40	·0250 0000 0000	4·2785 4303 8936	·0006 2500 0000	1·6202 4396 3007
41	·0243 9024 3902	4·3029 3328 2839	·0005 9488 3998	1·6208 3884 7005−
42	·0238 0952 3810	4·3267 4280 6648	·0005 6689 3424	1·6214 0574 0429
43	·0232 5581 3953	4·3499 9862 0602	·0005 4083 2883	1·6219 4657 3311
44	·0227 2727 2727	4·3727 2589 3329	·0005 1652 8926	1·6224 6310 2237
45	·0222 2222 2222	4·3949 4811 5551	·0004 9382 7160+	1·6229 5692 9397
46	·0217 3913 0435−	4·4166 8724 5986	·0004 7258 9792	1·6234 2951 9189
47	·0212 7659 5745−	4·4379 6384 1731	·0004 5269 3526	1·6238 8221 2716
48	·0208 3333 3333	4·4587 9717 5064	·0004 3402 7778	1·6243 1624 0494
49	·0204 0816 3265+	4·4792 0533 8329	·0004 1649 3128	1·6247 3273 3622
50	·0200 0000 0000	4·4992 0533 8329	·0004 0000 0000	1·6251 3273 3622

TABLE I (*continued*).

	Δ	Dibeta	Δ	Tribeta
n	$1/n$	$S(1/n)$	$1/n^2$	$S(1/n^2)$
51	·0196 0784 3137	4·5188 1318 1467	·0003 8446 7512	1·6255 1720 1134
52	·0192 3076 9231	4·5380 4395 0697	·0003 6982 2485+	1·6258 8702 3619
53	·0188 6792 4528	4·5569 1187 5226	·0003 5599 8576	1·6262 4302 2195+
54	·0185 1851 8519	4·5754 3039 3744	·0003 4293 5528	1·6265 8595 7723
55	·0181 8181 8182	4·5936 1221 1926	·0003 3057 8512	1·6269 1653 6236
56	·0178 5714 2857	4·6114 6935 4783	·0003 1887 7551	1·6272 3541 3787
57	·0175 4385 9649	4·6290 1321 4432	·0003 0778 7011	1·6275 4320 0798
58	·0172 4137 9310	4·6462 5459 3743	·0002 9726 5161	1·6278 4046 5959
59	·0169 4915 2542	4·6632 0374 6285+	·0002 8727 3772	1·6281 2773 9731
60	·0166 6666 6667	4·6798 7041 2952	·0002 7777 7778	1·6284 0551 7508
61	·0163 9344 2623	4·6962 6385 5575−	·0002 6874 4961	1·6286 7426 2469
62	·0161 2903 2258	4·7123 9288 7833	·0002 6014 5682	1·6289 3440 8151
63	·0158 7301 5873	4·7282 6590 3706	·0002 5195 2633	1·6291 8636 0784
64	·0156 2500 0000	4·7438 9090 3706	·0002 4414 0625	1·6294 3050 1409
65	·0153 8461 5385−	4·7592 7551 9090	·0002 3668 6391	1·6296 6718 7799
66	·0151 5151 5152	4·7744 2703 4242	·0002 2956 8411	1·6298 9675 6211
67	·0149 2537 3134	4·7893 5240 7376	·0002 2276 6763	1·6301 1952 2974
68	·0147 0588 2353	4·8040 5828 9729	·0002 1626 2976	1·6303 3578 5950−
69	·0144 9275 3623	4·8185 5104 3352	·0002 1003 9908	1·6305 4582 5857
70	·0142 8571 4286	4·8328 3675 7638	·0002 0408 1633	1·6307 4990 7490
71	·0140 8450 7042	4·8469 2126 4680	·0001 9837 3339	1·6309 4828 0829
72	·0138 8888 8889	4·8608 1015 3569	·0001 9290 1235−	1·6311 4118 2063
73	·0136 9863 0137	4·8745 0878 3706	·0001 8765 2468	1·6313 2883 4531
74	·0135 1351 3514	4·8880 2229 7220	·0001 8261 5047	1·6315 1144 9578
75	·0133 3333 3333	4·9013 5563 0553	·0001 7777 7778	1·6316 8922 7356
76	·0131 5789 4737	4·9145 1352 5290	·0001 7313 0194	1·6318 6235 7550+
77	·0129 8701 2987	4·9275 0053 8277	·0001 6866 2506	1·6320 3102 0056
78	·0128 2051 2821	4·9403 2105 1097	·0001 6436 5549	1·6321 9538 5605+
79	·0126 5822 7848	4·9529 7927 8946	·0001 6023 0732	1·6323 5561 6338
80	·0125 0000 0000	4·9654 7927 8946	·0001 5625 0000	1·6325 1186 6338
81	·0123 4567 9012	4·9778 2495 7958	·0001 5241 5790+	1·6326 6428 2128
82	·0121 9512 1951	4·9900 2007 9909	·0001 4872 0999	1·6328 1300 3127
83	·0120 4819 2771	5·0020 6827 2680	·0001 4515 8949	1·6329 5816 2076
84	·0119 0476 1905−	5·0139 7303 4585−	·0001 4172 3356	1·6330 9988 5432
85	·0117 6470 5882	5·0257 3774 0467	·0001 3840 8304	1·6332 3829 3736
86	·0116 2790 6977	5·0373 6564 7444	·0001 3520 8221	1·6333 7350 1957
87	·0114 9425 2874	5·0488 5990 0318	·0001 3211 7849	1·6335 0561 9807
88	·0113 6363 6364	5·0602 2353 6681	·0001 2913 2231	1·6336 3475 2038
89	·0112 3595 5056	5·0714 5949 1737	·0001 2624 6686	1·6337 6099 8724
90	·0111 1111 1111	5·0825 7060 2849	·0001 2345 6790	1·6338 8445 5514
91	·0109 8901 0989	5·0935 5961 3838	·0001 2075 8363	1·6340 0521 3877
92	·0108 6956 5217	5·1044 2917 9055−	·0001 1814 7448	1·6341 2336 1325−
93	·0107 5268 8172	5·1151 8186 7227	·0001 1562 0303	1·6342 3898 1628
94	·0106 3829 7872	5·1258 2016 5099	·0001 1317 3382	1·6343 5215 5009
95	·0105 2631 5789	5·1363 4648 0889	·0001 1080 3324	1·6344 6295 8333
96	·0104 1666 6667	5·1467 6314 7555+	·0001 0850 6944	1·6345 7146 5278
97	·0103 0927 8351	5·1570 7242 5906	·0001 0628 1220	1·6346 7774 6498
98	·0102 0408 1633	5·1672 7650 7539	·0001 0412 3282	1·6347 8186 9780
99	·0101 0101 0101	5·1773 7751 7640	·0001 0203 0405+	1·6348 8390 0185−
100	·0100 0000 0000	5·1873 7751 7640	·0001 0000 0000	1·6349 8390 0185−

TABLE II.

Table for finding the Derived Beta and Gamma Functions.

	Δ	Tetrabeta	Δ	Pentabeta
n	$1/n^3$	$S(1/n^3)$	$1/n^4$	$S(1/n^4)$
1	1·0000 0000 0000	1·0000 0000 0000	1·0000 0000 0000	1·0000 0000 0000
2	·1250 0000 0000	1·1250 0000 0000	·0625 0000 0000	1·0625 0000 0000
3	·0370 3703 7037	1·1620 3703 7037	·0123 4567 9012	1·0748 4567 9012
4	·0156 2500 0000	1·1776 6203 7037	·0039 0625 0000	1·0787 5192 9012
5	·0080 0000 0000	1·1856 6203 7037	·0016 0000 0000	1·0803 5192 9012
6	·0046 2962 9630−	1·1902 9166 6667	·0007 7160 4938	1·0811 2353 3951
7	·0029 1545 1895+	1·1932 0711 8562	·0004 1649 3128	1·0815 4002 7078
8	·0019 5312 5000	1·1951 6024 3562	·0002 4414 0625e	1·0817 8416 7703
9	·0013 7174 2112	1·1965 3198 5674	·0001 5241 5790+	1·0819 3658 3494
10	·0010 0000 0000	1·1975 3198 5674	·0001 0000 0000	1·0820 3658 3494
11	·0007 5131 4801	1·1982 8330 0475+	·0000 6830 1346	1·0821 0488 4839
12	·0005 7870 3704	1·1988 6200 4179	·0000 4822 5309	1·0821 5311 0148
13	·0004 5516 6136	1·1993 1717 0314	·0000 3501 2780−	1·0821 8812 2928
14	·0003 6443 1487	1·1996 8160 1801	·0000 2603 0820+	1·0822 1415 3748
15	·0002 9629 6296	1·1999 7789 8098	·0000 1975 3086	1·0822 3390 6835−
16	·0002 4414 0625e	1·2002 2203 8723	·0000 1525 8789	1·0822 4916 5624
17	·0002 0354 1624	1·2004 2558 0347	·0000 1197 3037	1·0822 6113 8660+
18	·0001 7146 7764	1·2005 9704 8111	·0000 0952 5987	1·0822 7066 4647
19	·0001 4579 3847	1·2007 4284 1958	·0000 0767 3360+	1·0822 7833 8008
20	·0001 2500 0000	1·2008 6784 1958	·0000 0625 0000	1·0822 8458 8008
21	·0001 0797 9700−	1·2009 7582 1658	·0000 0514 1890+	1·0822 8972 9898
22	·0000 9391 4350+	1·2010 6973 6008	·0000 0426 8834	1·0822 9399 8732
23	·0000 8218 9529	1·2011 5192 5537	·0000 0357 3458	1·0822 9757 2190−
24	·0000 7233 7963	1·2012 2426 3500+	·0000 0301 4082	1·0823 0058 6272
25	·0000 6400 0000	1·2012 8826 3500+	·0000 0256 0000	1·0823 0314 6272
26	·0000 5689 5767	1·2013 4515 9267	·0000 0218 8299	1·0823 0533 4570+
27	·0000 5080 5263	1·2013 9596 4531	·0000 0188 1676	1·0823 0721 6247
28	·0000 4555 3936	1·2014 4151 8467	·0000 0162 6926	1·0823 0884 3173
29	·0000 4100 2091	1·2014 8252 0558	·0000 0141 3865+	1·0823 1025 7038
30	·0000 3703 7037	1·2015 1955 7595−	·0000 0123 4568	1·0823 1149 1606
31	·0000 3356 7185−	1·2015 5312 4779	·0000 0108 2812	1·0823 1257 4419
32	·0000 3051 7578	1·2015 8364 2358	·0000 0095 3674	1·0823 1352 8093
33	·0000 2782 6474	1·2016 1146 8832	·0000 0084 3226	1·0823 1437 1319
34	·0000 2544 2703	1·2016 3691 1535−	·0000 0074 8315−	1·0823 1511 9634
35	·0000 2332 3615+	1·2016 6023 5150−	·0000 0066 6389	1·0823 1578 6023
36	·0000 2143 3471	1·2016 8166 8620	·0000 0059 5374	1·0823 1638 1397
37	·0000 1974 2167	1·2017 0141 0788	·0000 0053 3572	1·0823 1691 4970−
38	·0000 1822 4231	1·2017 1963 5019	·0000 0047 9585	1·0823 1739 4555−
39	·0000 1685 8005+	1·2017 3649 3024	·0000 0043 2257	1·0823 1782 6811
40	·0000 1562 5000	1·2017 5211 8024	·0000 0039 0625e	1·0823 1821 7436
41	·0000 1450 9366	1·2017 6662 7389	·0000 0035 3887	1·0823 1857 1323
42	·0000 1349 7462	1·2017 8012 4852	·0000 0032 1368	1·0823 1889 2691
43	·0000 1257 7509	1·2017 9270 2361	·0000 0029 2500+	1·0823 1918 5191
44	·0000 1173 9294	1·2018 0444 1655−	·0000 0026 6802	1·0823 1945 1994
45	·0000 1097 3937	1·2018 1541 5592	·0000 0024 3865+	1·0823 1969 5859
46	·0000 1027 3691	1·2018 2568 9283	·0000 0022 3341	1·0823 1991 9200−
47	·0000 0963 1777	1·2018 3532 1060	·0000 0020 4931	1·0823 2012 4131
48	·0000 0904 2245+	1·2018 4436 3305+	·0000 0018 8380	1·0823 2031 2511
49	·0000 0849 9860−	1·2018 5286 3165−	·0000 0017 3467	1·0823 2048 5978
50	·0000 0800 0000	1·2018 6086 3165−	·0000 0016 0000	1·0823 2064 5978

TABLE II (*continued*).

n	Δ $1/n^3$	Tetrabeta $S(1/n^3)$	Δ $1/n^4$	Pentabeta $S(1/n^4)$
51	·0000 0753 8579	1·2018 6840 1744	·0000 0014 7815+	1·0823 2079 3793
52	·0000 0711 1971	1·2018 7551 3714	·0000 0013 6769	1·0823 2093 0562
53	·0000 0671 6954	1·2018 8223 0669	·0000 0012 6735−	1·0823 2105 7297
54	·0000 0635 0658	1·2018 8858 1327	·0000 0011 7605−	1·0823 2117 4902
55	·0000 0601 0518	1·2018 9459 1845+	·0000 0010 9282	1·0823 2128 4184
56	·0000 0569 4242	1·2019 0028 6087	·0000 0010 1683	1·0823 2138 5867
57	·0000 0539 9773	1·2019 0568 5859	·0000 0009 4733	1·0823 2148 0600−
58	·0000 0512 5261	1·2019 1081 1121	·0000 0008 8367	1·0823 2156 8966
59	·0000 0486 9047	1·2019 1568 0168	·0000 0008 2526	1·0823 2165 1492
60	·0000 0462 9630−	1·2019 2030 9797	·0000 0007 7160+	1·0823 2172 8653
61	·0000 0440 5655+	1·2019 2471 5452	·0000 0007 2224	1·0823 2180 0877
62	·0000 0419 5898	1·2019 2891 1350+	·0000 0006 7676	1·0823 2186 8553
63	·0000 0399 9248	1·2019 3291 0599	·0000 0006 3480+	1·0823 2193 2033
64	·0000 0381 4697	1·2019 3672 5296	·0000 0005 9605−	1·0823 2199 1637
65	·0000 0364 1329	1·2019 4036 6625−	·0000 0005 6020+	1·0823 2204 7658
66	·0000 0347 8309	1·2019 4384 4934	·0000 0005 2702	1·0823 2210 0359
67	·0000 0332 4877	1·2019 4716 9811	·0000 0004 9625+	1·0823 2214 9984
68	·0000 0318 0338	1·2019 5035 0149	·0000 0004 6770−	1·0823 2219 6754
69	·0000 0304 4057	1·2019 5339 4206	·0000 0004 4117	1·0823 2224 0871
70	·0000 0291 5452	1·2019 5630 9658	·0000 0004 1649	1·0823 2228 2520+
71	·0000 0279 3991	1·2019 5910 3648	·0000 0003 9352	1·0823 2232 1872
72	·0000 0267 9184	1·2019 6178 2832	·0000 0003 7211	1·0823 2235 9083
73	·0000 0257 0582	1·2019 6435 3414	·0000 0003 5213	1·0823 2239 4297
74	·0000 0246 7771	1·2019 6682 1185−	·0000 0003 3348	1·0823 2242 7645−
75	·0000 0237 0370	1·2019 6919 1555+	·0000 0003 1605−	1·0823 2245 9250−
76	·0000 0227 8029	1·2019 7146 9584	·0000 0002 9974	1·0823 2248 9224
77	·0000 0219 0422	1·2019 7366 0006	·0000 0002 8447	1·0823 2251 7671
78	·0000 0210 7251	1·2019 7576 7257	·0000 0002 7016	1·0823 2254 4687
79	·0000 0202 8237	1·2019 7779 5494	·0000 0002 5674	1·0823 2257 0361
80	·0000 0195 3125	1·2019 7974 8619	·0000 0002 4414	1·0823 2259 4775−
81	·0000 0188 1676	1·2019 8163 0295+	·0000 0002 3231	1·0823 2261 8005+
82	·0000 0181 3671	1·2019 8344 3966	·0000 0002 2118	1·0823 2264 0123
83	·0000 0174 8903	1·2019 8519 2869	·0000 0002 1071	1·0823 2266 1194
84	·0000 0168 7183	1·2019 8688 0052	·0000 0002 0086	1·0823 2268 1280−
85	·0000 0162 8333	1·2019 8850 8385−	·0000 0001 9157	1·0823 2270 0437
86	·0000 0157 2189	1·2019 9008 0573	·0000 0001 8281	1·0823 2271 8718
87	·0000 0151 8596	1·2019 9159 9169	·0000 0001 7455+	1·0823 2273 6173
88	·0000 0146 7412	1·2019 9306 6581	·0000 0001 6675+	1·0823 2275 2848
89	·0000 0141 8502	1·2019 9448 5083	·0000 0001 5938	1·0823 2276 8787
90	·0000 0137 1742	1·2019 9585 6825+	·0000 0001 5242	1·0823 2278 4028
91	·0000 0132 7015−	1·2019 9718 3840	·0000 0001 4583	1·0823 2279 8611
92	·0000 0128 4211	1·2019 9846 8052	·0000 0001 3959	1·0823 2281 2570−
93	·0000 0124 3229	1·2019 9971 1281	·0000 0001 3368	1·0823 2282 5938
94	·0000 0120 3972	1·2020 0091 5253	·0000 0001 2808	1·0823 2283 8746
95	·0000 0116 6351	1·2020 0208 1604	·0000 0001 2277	1·0823 2285 1023
96	·0000 0113 0281	1·2020 0321 1884	·0000 0001 1774	1·0823 2286 2797
97	·0000 0109 5683	1·2020 0430 7567	·0000 0001 1296	1·0823 2287 4093
98	·0000 0106 2482	1·2020 0537 0050−	·0000 0001 0842	1·0823 2288 4934
99	·0000 0103 0610+	1·2020 0640 0660	·0000 0001 0410	1·0823 2289 5344
100	·0000 0100 0000	1·2020 0740 0660	·0000 0001 0000	1·0823 2290 5344

(3) *Moments of the Rank-Variates in Samples from the Exponential Curve.*

The equation to the exponential curve being

$$y = \frac{N}{\sigma} e^{-\frac{x}{\sigma}} \quad \text{.......................(i)},$$

if we measure the abscissae in terms of σ as unit, and take α_x to be the proportional area from the ordinate y at x to the tail, or

$$\alpha_x = \int_x^\infty y\,dx/N = e^{-x} \quad \text{.......................(ii)},$$

we have $$x = -\log \alpha_x.$$

Then the kth moment coefficient μ_k' of the character of the individual in the qth rank about the origin $x = 0$ is given by

$$\begin{aligned}\mu_k' &= \frac{1}{\mathrm{B}(q, n-q+1)} \int_0^1 (1-\alpha_x)^{q-1} \alpha_x^{n-q} x^k \, d\alpha_x \\ &= \frac{(-1)^k}{\mathrm{B}(q, n-q+1)} \int_0^1 (1-\alpha_x)^{q-1} \alpha_x^{n-q} (\log \alpha_x)^k \, d\alpha_x \\ &= \frac{(-1)^k}{\mathrm{B}(q, n-q+1)} \frac{d^k}{dn^k} \int_0^1 (1-\alpha_x)^{q-1} \alpha_x^{n-q} \, d\alpha_x \\ &= \frac{(-1)^k}{\mathrm{B}(q, n-q+1)} \frac{d^k}{dn^k} \mathrm{B}(q, n-q+1) \quad \text{.......................(iii)}.\end{aligned}$$

We will write $\mathfrak{B}(q, n)$ for this B-function or where there is no danger of confusion simply $\mathfrak{B}$. The next task is to transfer these moment coefficients to the mean by the aid of the well-known formula

$$\mu_k = \mu_k' - k\mu'_{k-1}\mu_1' + \frac{k(k-1)}{2!} \mu'_{k-2} \mu_1'^2 - \dots.$$

From this we easily deduce by (iii) that

$$\mu_k = \frac{(-1)^k}{\mathfrak{B}} \left(\frac{d}{dn} + \mu_1'\right)^k \mathfrak{B} \quad \text{.......................(iv)}.$$

Now the operator $\dfrac{d}{dn} + \mu_1^2$ applied to $\mathfrak{B}$ gives the result $\mathfrak{B}\left(\dfrac{d \log \mathfrak{B}}{dn} + \mu_1'\right)$.

Hence $$\mu_1 = -\left(\frac{d \log \mathfrak{B}}{dn} + \mu_1'\right) = 0 \quad \text{.......................(v)}.$$

Thus the mean character ${}_n\bar{x}_q$ in the qth rank in samples of n from an exponential curve is given,

$${}_n\bar{x}_q = \mu_1' = -\frac{d \log \mathfrak{B}}{dn} \quad \text{.......................(vi)},$$

or reinserting σ, $${}_n\bar{x}_q = \sigma\left(\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-q+1}\right) \quad \text{.......................(vi}^{\text{bis}}\text{)}.$$

Again, $$\begin{aligned}\mu_2 &= \frac{1}{\mathfrak{B}}\left(\frac{d}{dn} + \mu_1\right) \mathfrak{B}\left(\frac{d \log \mathfrak{B}}{dn} + \mu_1'\right) \\ &= \left(\frac{d \log \mathfrak{B}}{dn} + \mu_1'\right)^2 + \frac{d^2 \log \mathfrak{B}}{dn^2}.\end{aligned}$$

The squared factor vanishes by (v) and we have

$$ {}_n\sigma_q^{\,2} = \mu_2 = \frac{d^2 \log \mathfrak{F}}{dn^2} \quad \text{.............................(vii),}$$

or reinserting σ^2,

$$ {}_n\sigma_q^{\,2} = \sigma^2 \left(\frac{1}{n^2} + \frac{1}{(n-1)^2} + \dots + \frac{1}{(n-q+1)^2}\right) \quad \text{...........(vii}^{\text{bis}}\text{).}$$

Proceeding to the third moment by applying the operator $\frac{d}{dn} + \mu_1'$ once more, we have

$$\mu_3 = -\frac{1}{\mathfrak{F}}\left[\mathfrak{F}\left(\frac{d\log\mathfrak{F}}{dn} + \mu_1'\right)^3 + 3\mathfrak{F}\left(\frac{d\log\mathfrak{F}}{dn} + \mu_1'\right)\frac{d^2\log\mathfrak{F}}{dn^2} + \mathfrak{F}\frac{d^3\log\mathfrak{F}}{dn^3}\right]$$

$$= -\frac{d^3\log\mathfrak{F}}{dn^3},$$

or reinserting σ^3,

$$\mu_3 = -\sigma^3\frac{d^3\log\mathfrak{F}}{dn^3} \quad \text{.......................................(viii)}$$

$$= 2\sigma^3\left(\frac{1}{n^3} + \frac{1}{(n-1)^3} + \dots + \frac{1}{(n-q+1)^3}\right) \quad \text{...........(viii}^{\text{bis}}\text{).}$$

Again,

$$\mu_4 = \frac{1}{\mathfrak{F}}\left[\mathfrak{F}\left(\frac{d\log\mathfrak{F}}{dn} + \mu_1'\right)^4 + 6\mathfrak{F}\left(\frac{d\log\mathfrak{F}}{dn} + \mu_1'\right)^2\frac{d^2\log\mathfrak{F}}{dn^2} + 3\mathfrak{F}\left(\frac{d^2\log\mathfrak{F}}{dn^2}\right)^2 \right.$$

$$\left. + 4\mathfrak{F}\left(\frac{d\log\mathfrak{F}}{dn} + \mu_1'\right)\frac{d^3\log\mathfrak{F}}{dn^3} + \mathfrak{F}\frac{d^4\log\mathfrak{F}}{dn^4}\right],$$

or

$$\mu_4 = \frac{d^4\log\mathfrak{F}}{dn^4} + 3\left(\frac{d^2\log\mathfrak{F}}{dn^2}\right)^2 \quad \text{......................(ix),}$$

or reinserting σ^4,

$$\mu_4 = \sigma^4\left[6\left(\frac{1}{n^4} + \frac{1}{(n-1)^4} + \dots + \frac{1}{(n-q+1)^4}\right)\right.$$

$$\left. + 3\left(\frac{1}{n^2} + \frac{1}{(n-1)^2} + \dots + \frac{1}{(n-q+1)^2}\right)^2\right] \quad \text{...........(ix}^{\text{bis}}\text{).}$$

Similarly,

$$\mu_5 = -\left(10\,\frac{d^2\log\mathfrak{F}}{dn^2}\,\frac{d^3\log\mathfrak{F}}{dn^3} + \frac{d^5\log\mathfrak{F}}{dn^5}\right) \quad \text{.......................................(x),}$$

$$= \sigma^5\left[20\left(\frac{1}{n^2} + \frac{1}{(n-1)^2} + \dots + \frac{1}{(n-q+1)^2}\right)\left(\frac{1}{n^3} + \frac{1}{(n-1)^3} + \dots + \frac{1}{(n-q+1)^3}\right)\right.$$

$$\left. + 24\left(\frac{1}{n^5} + \frac{1}{(n-1)^5} + \dots + \frac{1}{(n-q+1)^5}\right)\right] \quad \text{.........(x}^{\text{bis}}\text{).}$$

Still higher moments can be found readily by repeated use of the operator $\frac{d}{dn} + \mu_1'$. It is clear that such moment coefficients are all expressible in terms of the hyperbeta functions and these in terms of the series of the inverse powers of integer numbers.

I have succeeded in finding the analytical form of the frequency distribution of the variate of the individual in the qth rank; this will be given later. Since each additional moment involves the sum of a series of one degree higher inverse powers,

it was most unlikely that a curve with a linear relation between successive moments would theoretically describe this frequency; it would have involved a relation between series of inverse powers of a very improbable character. In actual practice, however, I find that we get an adequate representation of the frequency by using a Pearson curve found from ${}_n\sigma_q$, and the corresponding β_1 and β_2. These latter are

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{4\left(\frac{1}{n^3} + \frac{1}{(n-1)^3} + \ldots + \frac{1}{(n-q+1)^3}\right)^2}{\left(\frac{1}{n^2} + \frac{1}{(n-1)^2} + \ldots + \frac{1}{(n-q+1)^2}\right)^3} \quad \ldots\ldots\ldots\ldots\text{(xi)},$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{6\left(\frac{1}{n^4} + \frac{1}{(n-1)^4} + \ldots + \frac{1}{(n-q+1)^4}\right)}{\left(\frac{1}{n^2} + \frac{1}{(n-1)^2} + \ldots + \frac{1}{(n-q+1)^2}\right)^2} \quad \ldots\ldots\ldots\text{(xii)}.$$

A particular case is easy. Suppose q to represent the first individual. Then

$${}_n\sigma_q = \sigma/n, \quad \beta_1 = 4, \quad \beta_2 = 9, \quad \mu_5 = \sigma^5\frac{44}{n^5}, \text{ etc.},$$

but these are the constants belonging to an exponential curve of standard deviation σ/n. This result has already been obtained by E. S. Pearson and J. Neyman*.

Let us now consider the last individual in a sample of 100, or $q = 100$; then we easily obtain, from our Tables I and II,

$${}_{100}\bar{x}_{100} = \sigma \times 5{\cdot}187{,}3775^+,$$
$${}_{100}\sigma_{100} = \sigma \times 1{\cdot}278{,}665^-,$$
$$\beta_1 = 1{\cdot}32231, \quad \beta_2 = 5{\cdot}59593.$$

Let us now consider an individual near the median, $q = 50$, say, in a sample of 100. Here

$${}_{100}\bar{x}_{50} = \sigma\ (5{\cdot}187{,}3775 - 4{\cdot}499{,}2053) = \sigma \times {\cdot}688{,}1722,$$
$${}_{100}\sigma_{50}^2 = \sigma^2(1{\cdot}634{,}9839 - 1{\cdot}625{,}1327) = \sigma^2({\cdot}009{,}8512),$$

and
$${}_{100}\sigma_{50} = \sigma \times {\cdot}992{,}530,$$
$$\beta_1 = {\cdot}089{,}845, \quad \beta_2 = 3{\cdot}139{,}689.$$

Remembering that for the first rank

$$\beta_1 = 4 \text{ and } \beta_2 = 9,$$

we see that even for a sample of 100 the distribution of extreme ranks differs much from normality, the last ranks being more nearly normal than the early ranks, while ranks in the neighbourhood of the median are approaching normality but still differ from it both in skewness and kurtosis. It is clear that the approach to a normal distribution depends not only on the size of the sample, but very largely on the rank of the individual in the sample. It would be an interesting task to trace for different values of q the family of curves β_1, β_2 as n is varied. Each member of the family appears to run through several Pearson-types.

* *Biometrika*, Vol. xx^A, p. 223. The reader will find in that paper another method of treating samples from the exponential curve (pp. 221—30).

(4) *Product Moments of Rank-Variates in Samples from the Exponential Curve.*

We next turn to the mean product of the variates of two ranked characters x_q, $x_{q'}$ in a sample of size n. If as before we measure the area ratio from x to α we have

$$x = -\log \alpha_x,$$

and

$$\{x_q{}^t x_{q'}{}^s\} = (-1)^{s+t} \chi \int_0^1 d\alpha_x' \alpha_x'^{n-q} (\log \alpha_x')^s \int_{\alpha_x'}^1 d\alpha_x (1-\alpha_x)^{q-1} (\alpha_x - \alpha_x')^{q'-q-1} (\log \alpha_x)^t \quad \text{.........(xiii)},$$

where $x_{q'} > x_q$ and $\alpha_x' < \alpha_x$, and

$$\chi = \frac{\Gamma(n+1)\,\sigma^{s+t}}{\Gamma(q)\,\Gamma(q'-q)\,\Gamma(n-q'+1)}.$$

We have
$$\{x_q{}^t x_{q'}{}^s\} = (-1)^{s+t} \chi \frac{d^s}{dn^s} \int_0^1 d\alpha_x' \alpha_x'^{n-q} u_{q'-q-1},$$

where
$$u_{q'-q-1} = \int_{\alpha_x'}^1 d\alpha_x (1-\alpha_x)^{q-1} (\alpha_x - \alpha_x')^{q'-q-1} (\log \alpha_x)^t.$$

Now integrate by parts and note that $\alpha_x' = 0$ at the lower limit and $u_{q'-q-1} = 0$ at the upper limit. Hence

$$\{x_q{}^t x_{q'}{}^s\} = (-1)^{s+t} \chi \frac{d^s}{dn^s} \left(\frac{1}{n-q+1} \int_0^1 d\alpha_x' \alpha_x'^{n-q+1} (-1) \frac{du_{q'-q-1}}{d\alpha_x'} \right).$$

But
$$\frac{du_{q'-q-1}}{d\alpha_x'} = [-(1-\alpha_x)^{q-1} (\alpha_x - \alpha_x')^{q'-q-1} (\log \alpha_x)^t]_{\alpha_x = \alpha_x'}$$
$$- (q'-q-1) \int_{\alpha_x'}^1 d\alpha_x (1-\alpha_x)^{q-1} (\alpha_x - \alpha_x')^{q'-q-2} (\log \alpha_x)^t,$$

and the term between square brackets vanishing,

$$\{x_q{}^t x_{q'}{}^s\} = (-1)^{s+t} \chi \frac{d^s}{dn^s} \frac{q'-q-1}{n-q+1} \int_0^1 d\alpha_x' \alpha_x'^{n-q+1} u_{q'-q-2}.$$

Now this process may be repeated until the whole power of $\alpha_x - \alpha_x'$ is transferred to α_x' and, when this is accomplished, we have

$$\{x_q{}^t x_{q'}{}^s\} = (-1)^{s+t} \chi \frac{d^s}{dn^s} \left(\frac{(q'-q-1)(q'-q-2)\ldots 1}{(n-q'+1)(n-q'+2)\ldots n-q-1} \right.$$
$$\left. \times \int_0^1 d\alpha_x' \alpha_x'^{n-q-1} \int_{\alpha_x'}^1 d\alpha_x (1-\alpha_x)^{q-1} (\log \alpha_x)^t \right) \quad \text{.........(xiv)}.$$

The double integral may be expressed as

$$\int_0^1 u_0 \frac{d(\alpha_x'^{n-q})}{n-q} = \frac{1}{n-q} \left[(\alpha_x'^{n-q} u_0)_0^1 - \int_0^1 \alpha_x'^{n-q} \frac{du_0}{d\alpha_x'} d\alpha_x' \right],$$

where the part outside the integration vanishes and the integral

$$\int_0^1 \alpha_x'^{n-q} \frac{du_0}{d\alpha_x'} d\alpha_x' = -\int_0^1 \alpha_x'^{n-q} (1-\alpha_x')^{q-1} (\log \alpha_x')^t d\alpha_x'$$
$$= -\frac{d^t}{dn^t} \mathrm{B}(q, n-q+1).$$

We can accordingly write (xiv) in the form

$$\{x_q{}^t x_{q'}{}^s\} = (-1)^{s+t}\sigma^{s+t}\frac{\Gamma(n+1)}{\Gamma(q'-q)\,\Gamma(q)\,\Gamma(n-q'+1)}\frac{d^s}{dn^s}$$
$$\left(\frac{\Gamma(q'-q)\,\Gamma(n-q'+1)}{\Gamma(n-q+1)}\times\frac{d^t}{dn^t}\mathrm{B}(q,\,n-q+1)\right)$$
$$=(-1)^{s+t}\sigma^{s+t}\frac{1}{\mathrm{B}(q'-q,\,n-q'+1)\,\mathrm{B}(q,\,n-q+1)}\frac{d^s}{dn^s}$$
$$\times\left(\mathrm{B}(q'-q,\,n-q'+1)\frac{d^t}{dn^t}\mathrm{B}(q,\,n-q+1)\right)\ \ldots\text{(xv)}.$$

The product moments of the various orders are taken about the origin of x, and must be reduced to the mean by the usual process. The formula (xv) contains all the results previously reached. If we write $s=0$, $t=k$, we have

$$\{\mu'_q{}^k\} = \frac{(-1)^k\sigma^k}{\mathrm{B}(q,\,n-q+1)}\frac{d^k\,\mathrm{B}(q,\,n-q+1)}{dn^k},$$ see p. 210, Eqn. (iii).

If we put $t=0$, $s=k$, we obtain, after some slight reductions,

$$\{\mu'_{q'}{}^k\} = \frac{(-1)^k\sigma^k}{\mathrm{B}(q',\,n-q'+1)}\frac{d^k\,\mathrm{B}(q',\,n-q'+1)}{dn^k}.$$

If we take $s=1$, $t=1$, we have

$$\{x_q x_{q'}\} = \sigma^2\frac{1}{\mathrm{B}(q'-q,\,n-q'+1)\,\mathrm{B}(q,\,n-q+1)}\left(\frac{d\,\mathrm{B}(q'-q,\,n-q'+1)}{dn}\right.$$
$$\left.\times\frac{d\,\mathrm{B}(q,\,n-q+1)}{dn}+\mathrm{B}(q'-q,\,n-q'+1)\frac{d^2\mathrm{B}(q,\,n-q+1)}{dn^2}\right)$$
$$=\sigma^2\left(\frac{d\log\mathrm{B}(q'-q,\,n-q'+1)}{dn}\frac{d\log\mathrm{B}(q,\,n-q+1)}{dn}\right.$$
$$\left.+\frac{1}{\mathrm{B}(q,\,n-q+1)}\frac{d^2\mathrm{B}(q,\,n-q+1)}{dn^2}\right).$$

Now

$$\frac{d\log\mathrm{B}(q'-q,\,n-q'+1)}{dn} = \frac{\Gamma'(n-q'+1)}{\Gamma(n-q'+1)}-\frac{\Gamma'(n-q+1)}{\Gamma(n-q+1)}$$
$$=\frac{\Gamma'(n-q'+1)}{\Gamma(n-q'+1)}-\frac{\Gamma'(n+1)}{\Gamma(n+1)}-\left(\frac{\Gamma'(n-q+1)}{\Gamma(n-q+1)}-\frac{\Gamma'(n+1)}{\Gamma(n+1)}\right)$$
$$=\frac{d\log\mathrm{B}(q',\,n-q'+1)}{dn}-\frac{d\log\mathrm{B}(q,\,n-q+1)}{dn}$$
$$=-\frac{\bar{x}_{q'}}{\sigma}+\frac{\bar{x}_q}{\sigma}.$$

Thus $$\{x_q x_{q'}\} = \bar{x}_q\bar{x}_{q'}-\bar{x}_q{}^2+\frac{1}{\mathrm{B}(q,\,n-q+1)}\frac{d^2\mathrm{B}(q,\,n-q+1)}{dn^2},$$

or $$\{x_q x_{q'}\}-\bar{x}_q\bar{x}_{q'} = {}_q\mu_2'-\bar{x}_q{}^2=\sigma_q{}^2\ \ldots\ldots\ldots\ldots\ldots\text{(xvi)},$$

and accordingly $$r_{x_q x_{q'}} = \frac{\{x_q x_{q'}\}-\bar{x}_q\bar{x}_{q'}}{\sigma_q\sigma_{q'}} = \frac{\sigma_q}{\sigma_{q'}}\ \ldots\ldots\ldots\ldots\ldots\text{(xvi}^{\text{bis}}\text{)},$$

a very simple formula for the correlation coefficient of x_q and $x_{q'}$, $q'>q$.

Now consider the interval between the qth and q'th ranks $= x_{q'} - x_q$ with a mean value $\bar{x}_q - \bar{x}_{q'}$. Then

$$\begin{aligned}\sigma^2_{x_{q'}-x_q} &= \{((x_{q'} - x_q) - (\bar{x}_{q'} - \bar{x}_q))^2\}\\ &= \sigma^2_{q'} + \sigma^2_q - 2r_{x_q x_{q'}}\sigma_q\sigma_{q'}\\ &= \sigma^2_{q'} - \sigma^2_q \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xvii)}.\end{aligned}$$

Let us next investigate the correlation coefficient between the intervals $x_{p'} - x_p$ with $x_{q'} - x_q$,

$$q' > q > p' > p.$$

The mean product is

$$\begin{aligned}&\{((x_{p'} - x_p) - (\bar{x}_{p'} - \bar{x}_p))\,((x_{q'} - x_q) - (\bar{x}_{q'} - \bar{x}_q))\}\\ &\quad = \{((x_{p'} - \bar{x}_{p'}) - (x_p - \bar{x}_p))\,((x_{q'} - \bar{x}_{q'}) - (x_q - \bar{x}_q))\}\\ &\quad = \{(x_{p'} - \bar{x}_{p'})(x_{q'} - \bar{x}_q) + (x_p - \bar{x}_p)(x_q - \bar{x}_q) - (x_p - \bar{x}_p)(x_{q'} - \bar{x}_{q'}) - (x_{p'} - \bar{x}_{p'})(x_q - \bar{x}_q)\}\\ &\quad = \sigma_{p'}\sigma_{q'}r_{x_{p'}x_{q'}} + \sigma_p\sigma_q r_{x_p x_q} - \sigma_p\sigma_{q'}r_{x_p x_{q'}} - \sigma_{p'}\sigma_q r_{x_{p'}x_q}\\ &\quad = \sigma^2_{p'} + \sigma^2_p - \sigma^2_p - \sigma^2_{p'} = 0,\end{aligned}$$

or, the correlation coefficient between any two non-covering intervals is zero.

Since

$$x_{q'} - x_q = (x_{q'} - x_{q'-1}) + (x_{q'-1} - x_{q'-2}) + (x_{q'-2} - x_{q'-3}) + \ldots + (x_{q+1} - x_q),$$

and

$$\bar{x}_{q'} - \bar{x}_q = (\bar{x}_{q'} - \bar{x}_{q'-1}) + (\bar{x}_{q'-1} - \bar{x}_{q'-2}) + (\bar{x}_{q'-2} - \bar{x}_{q'-3}) + \ldots + (\bar{x}_{q+1} - \bar{x}_q),$$

it follows by subtracting, squaring and taking mean values that

$$\begin{aligned}\sigma^2_{x_{q'}-x_q} &= \sigma^2_{x_{q'}-x_{q'-1}} + \sigma^2_{x_{q'-1}-x_{q'-2}} + \sigma^2_{x_{q'-2}-x_{q'-3}} + \ldots + \sigma^2_{x_{q+1}-x_q}\\ &\quad - \frac{1}{(n-q'+1)^2} + \frac{1}{(n-q'+2)^2} + \frac{1}{(n-q'+3)^2} + \ldots + \frac{1}{(n-q)^2}\\ &= \sigma^2_{q'} - \sigma^2_q, \text{ as before.}\end{aligned}$$

We pass now to higher product moments of the qth and q'th ranks* and to the higher moments of the interrank interval. For brevity we shall write

$$\mathrm{B}(q', n-q'+1) = \mathfrak{B}' \quad\text{and}\quad \mathrm{B}(q, n-q+1) = \mathfrak{B},$$

while the corresponding hyperbeta functions will be

$$\mathfrak{B}' = \frac{d}{dn}\log \mathrm{B}(q', n-q'+1), \quad \mathfrak{B} = \frac{d^2}{dn^2}\log \mathrm{B}(q, n-q+1),$$

and so on.

We note that

$$\frac{1}{\mathrm{B}(q, n-q+1)}\,\frac{d\,\mathrm{B}(q, n-q+1)}{dn} = \mathfrak{B} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xviii)},$$

$$\frac{1}{\mathrm{B}(q, n-q+1)}\,\frac{d^2\mathrm{B}(q, n-q+1)}{dn^2} = \mathfrak{B} + \mathfrak{B}^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(xix)},$$

$$\frac{1}{\mathrm{B}(q, n-q+1)}\,\frac{d^3\mathrm{B}(q, n-q+1)}{dn^3} = \mathfrak{B} + 3\mathfrak{B}\mathfrak{B} + \mathfrak{B}^3 \quad \ldots\ldots\ldots\ldots\text{(xx)},$$

$$\frac{1}{\mathrm{B}(q, n-q+1)}\,\frac{d^4\mathrm{B}(q, n-q+1)}{dn^4} = \mathfrak{B} + 3\mathfrak{B}^2 + 4\mathfrak{B}\mathfrak{B} + 6\mathfrak{B}^2\mathfrak{B} + \mathfrak{B}^4 \quad \ldots\text{(xxi)},$$

* We suppose throughout $q' > q$, i.e. $x_{q'} > x_q$.

while dropping the affix n, we have for the qth rank

$$\bar{x}_q = -\sigma\mathfrak{f}, \quad \sigma^2_q = \sigma^2\mathfrak{f}, \quad \mu_3(q) = -\sigma^3\mathfrak{f}, \quad \mu_4 = \sigma^4(\mathfrak{f} + 3\mathfrak{f}^2) \quad \ldots\text{(xxii)}.$$

Starting from (xiv) we can write it in the simpler form below by expanding the B-functions in Γ-functions

$$\{x_q{}^t x_{q'}{}^s\} = (-1)^{s+t}\sigma^{s+t}\frac{1}{\mathfrak{f}'}\frac{d^s}{dn^s}\left(\mathfrak{f}'\frac{1}{\mathfrak{f}}\frac{d^t\mathfrak{f}}{dn^t}\right) \quad \ldots\ldots\ldots\ldots\text{(xxiii)}.$$

Here by aid of equations (xv) to (xviii), or still further such equations, it is fairly easy to obtain any required product moment about the start of the exponential curve as origin.

The third order products are

(i) $t=1$, $s=2$:

$$\{x_q x^2_{q'}\} = -\sigma^3\frac{1}{\mathfrak{f}'}\frac{d^2}{dn^2}(\mathfrak{f}'\mathfrak{f})$$
$$= -\sigma^3\frac{1}{\mathfrak{f}'}\left(\mathfrak{f}\frac{d^2\mathfrak{f}'}{dn^2} + 2\frac{d\mathfrak{f}'}{dn}\mathfrak{f} + \mathfrak{f}'\mathfrak{f}\right),$$

and thus by (xviii) and (xix),

$$\{x_q x^2_{q'}\} = -\sigma^3(\mathfrak{f}\mathfrak{f}' + \mathfrak{f}\mathfrak{f}'^2 + 2\mathfrak{f}'\mathfrak{f} + \mathfrak{f}) \quad \ldots\ldots\ldots\ldots\text{(xxiv)},$$

or by aid of (xix),

$$\{x_q x^2_{q'}\} = \bar{x}_q\sigma^2_{q'} + \bar{x}_q\bar{x}^2_{q'} + 2\bar{x}_{q'}\sigma^2_q + \mu_3(q) \quad \ldots\ldots\ldots\text{(xxiv}^{\text{bis}}\text{)}.$$

(ii) $t=2$, $s=1$:

$$\{x^2_q x_{q'}\} = -\sigma^3\frac{1}{\mathfrak{f}'}\frac{d}{dn}\left(\mathfrak{f}'\frac{1}{\mathfrak{f}}\frac{d^2\mathfrak{f}}{dn^2}\right) = -\sigma^3\frac{1}{\mathfrak{f}'}\frac{d}{dn}(\mathfrak{f}'(\mathfrak{f} + \mathfrak{f}^2))$$
$$= -\sigma^3\left(\frac{1}{\mathfrak{f}'}\frac{d\mathfrak{f}'}{dn}(\mathfrak{f} + \mathfrak{f}^2) + \mathfrak{f} + 2\mathfrak{f}\mathfrak{f}\right)$$
$$= -\sigma^3(\mathfrak{f}'\mathfrak{f} + \mathfrak{f}\mathfrak{f}^2 + 2\mathfrak{f}\mathfrak{f} + \mathfrak{f}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(xxv)},$$

or, we may put it in the form

$$\{x^2_q x_{q'}\} = \bar{x}_{q'}\sigma^2_q + \bar{x}_{q'}\bar{x}^2_q + 2\bar{x}_q\sigma^2_q + \mu_3(q) \quad \ldots\ldots\ldots\ldots\text{(xxv}^{\text{bis}}\text{)}.$$

We are now in a position to find

$$\{(x_{q'} - x_q)^3\} = \{x^3_{q'}\} - 3\{x^2_{q'}x_q\} + 3\{x_{q'}x^2_q\} - \{x^3_q\}$$
$$= \mu_3'(q') + 3\bar{x}_{q'}\sigma^2_q - 3\bar{x}_q\sigma^2_{q'} + 3\bar{x}_{q'}\bar{x}^2_q - 3\bar{x}^2_{q'}\bar{x}_q$$
$$+ 6\bar{x}_q\sigma^2_q - 6\bar{x}_{q'}\sigma^2_q - \mu_3'(q) \ldots\ldots\text{(xxvi)},$$

where $\mu_3(q)$, $\mu_3(q')$, $\mu_3'(q)$, $\mu_3'(q')$ denote respectively the third moment coefficients of the qth and q'th ranks respectively about their means and about the start of the exponential curve.

It is now easy to obtain the third moment coefficient of the interval between the qth and q'th ranked variates about its mean $\bar{x}_{q'} - \bar{x}_q$.

$$\mu_3(x_{q'} - x_q) = \mu_{3\,(q'-q)}, \text{ say,}$$
$$= \{(x_{q'} - x_q - (\bar{x}_{q'} - \bar{x}_q))^3\}$$
$$= \{(x_{q'} - x_q)^3\} - 3\{(x_{q'} - x_q)^2\}(\bar{x}_{q'} - \bar{x}_q) + 2(\bar{x}_{q'} - \bar{x}_q)^3$$
$$= \{(x_{q'} - x_q)^3\} - 3\{(x_{q'} - x_q - (\bar{x}_{q'} - \bar{x}_q))^2\}(\bar{x}_{q'} - \bar{x}_q) - (\bar{x}_{q'} - \bar{x}_q)^3$$
$$= \{(x_{q'} - x_q)^3\} - 3\mu_{2\,(q'-q)}(\bar{x}_{q'} - \bar{x}_q) - (\bar{x}_{q'} - \bar{x}_q)^3.$$

Now substitute from (xxvi) and (xvii), and we have

$$\begin{aligned}\mu_{3\,(q'-q)} &= \mu_3'(q') + 3\bar{x}_{q'}\sigma^2_q - 3\bar{x}_q\sigma^2_{q'} + 3\bar{x}_{q'}\bar{x}^2_q - 3\bar{x}^2_{q'}\bar{x}_q + 6\bar{x}_q\sigma^2_q - 6\bar{x}_{q'}\sigma^2_q \\ &\qquad - \mu_3'(q) - 3(\sigma^2_{q'} - \sigma^2_q)(\bar{x}_{q'} - \bar{x}_q) - \bar{x}^3_{q'} + 3\bar{x}^2_{q'}\bar{x}_q - 3\bar{x}_{q'}\bar{x}^2_q + \bar{x}^3_q \\ &= \mu_3'(q') - 3\bar{x}_{q'}\sigma^2_{q'} - \bar{x}^3_{q'} - (\mu_3'(q) - 3\bar{x}_q\sigma^2_q - \bar{x}^3_q) \\ &= \mu_3'(q') - 3\bar{x}_{q'}\mu_2'(q') + 2\bar{x}^3_{q'} - (\mu_3'(q) - 3\bar{x}_q\mu_2'(q) + 2\bar{x}^3_q) \\ &= \mu_3(q') - \mu_3(q).\end{aligned}$$

Thus $$\mu_{3\,(q'-q)} = \sigma^3\left(\frac{d^3\log\mathfrak{F}}{dn^3} - \frac{d^3\log\mathfrak{F}'}{dn^3}\right) = \sigma^3(\mathfrak{F}' - \mathfrak{F}) \quad \text{.....................(xxvii).}$$

From this we deduce at once

$$\beta_1(q'-q) = \frac{(\mathfrak{F}' - \mathfrak{F})^2}{(\mathfrak{F}' - \mathfrak{F})^3} \quad \text{.....................(xxvii}^{\text{bis}}\text{).}$$

We must next consider the fourth order product moment coefficients with the view of determining $\beta_2(q'-q)$.

(iii) $t = 1,\ s = 3$:

$$\begin{aligned}\{x_q x^3_{q'}\} &= \sigma^4\frac{1}{\mathfrak{F}'}\frac{d^3}{dn^3}\left(\mathfrak{F}'\frac{1}{\mathfrak{F}}\frac{d\mathfrak{F}}{dn}\right) = \sigma^4\frac{1}{\mathfrak{F}'}\frac{d^3}{dn^3}(\mathfrak{F}'\mathfrak{F}) \\ &= \sigma^4\left(\frac{1}{\mathfrak{F}'}\frac{d^3\mathfrak{F}'}{dn^3}\mathfrak{F} + 3\frac{1}{\mathfrak{F}'}\frac{d^2\mathfrak{F}'}{dn^2}\mathfrak{F} + 3\frac{1}{\mathfrak{F}'}\frac{d\mathfrak{F}'}{dn}\mathfrak{F} + \mathfrak{F}\right) \\ &= \sigma^4((\mathfrak{F}' + 3\mathfrak{F}'\mathfrak{F}' + \mathfrak{F}'^3)\mathfrak{F} + 3(\mathfrak{F}' + \mathfrak{F}'^2)\mathfrak{F} + 3\mathfrak{F}'\mathfrak{F} + \mathfrak{F}) \quad \text{...........(xxviii),} \\ &= (\mu_3(q')\bar{x}_q + 3\sigma^2_{q'}\bar{x}_{q'}\bar{x}_q + \bar{x}^3_{q'}\bar{x}_q + 3\sigma^2_{q'}\sigma^2_q + 3\bar{x}^2_{q'}\sigma^2_q + 3\bar{x}_{q'}\mu_3(q) + \mu_4(q) - 3\sigma^4_q)\end{aligned}$$

......(xxviii$^{\text{bis}}$).

(iv) $t = 3,\ s = 1$:

$$\begin{aligned}\{x^3_q x_{q'}\} &= \sigma^4\frac{1}{\mathfrak{F}'}\frac{d}{dn}\left(\mathfrak{F}'\frac{1}{\mathfrak{F}}\frac{d^3\mathfrak{F}}{dn^3}\right) \\ &= \sigma^4\frac{1}{\mathfrak{F}'}\frac{d}{dn}(\mathfrak{F}'(\mathfrak{F} + 3\mathfrak{F}\mathfrak{F} + \mathfrak{F}^3)) \\ &= \sigma^4(\mathfrak{F}'(\mathfrak{F} + 3\mathfrak{F}\mathfrak{F} + \mathfrak{F}^3) + \mathfrak{F} + 3\mathfrak{F}^2 + 3\mathfrak{F}\mathfrak{F} + 3\mathfrak{F}^2\mathfrak{F}) \quad \text{.................(xxix),}\end{aligned}$$

or $$= \bar{x}_{q'}\mu_3(q) + 3\bar{x}_{q'}\bar{x}_q\sigma^2_q + \bar{x}_{q'}\bar{x}^3_q + \mu_4(q) - 3\sigma^4_q + 3\sigma^4_q + 3\bar{x}_q\mu_3(q) + 3\bar{x}^2_q\sigma^2_q$$

......(xxix$^{\text{bis}}$).

(v) $t = 2,\ s = 2$:

$$\begin{aligned}\{x^2_q x^2_{q'}\} &= \sigma^4\frac{1}{\mathfrak{F}'}\frac{d^2}{dn^2}\left(\mathfrak{F}'\frac{1}{\mathfrak{F}}\frac{d^2\mathfrak{F}}{dn^2}\right) = \sigma^4\frac{1}{\mathfrak{F}'}\frac{d^2}{dn^2}\mathfrak{F}'(\mathfrak{F} + \mathfrak{F}^2) \\ &= \sigma^4\frac{1}{\mathfrak{F}'}\left(\frac{d^2\mathfrak{F}'}{dn^2}(\mathfrak{F} + \mathfrak{F}^2) + 2\frac{d\mathfrak{F}'}{dn}(\mathfrak{F} + 2\mathfrak{F}\mathfrak{F}) + \mathfrak{F}'(\mathfrak{F} + 2\mathfrak{F}^2 + 2\mathfrak{F}\mathfrak{F})\right) \\ &= \sigma^4((\mathfrak{F}' + \mathfrak{F}'^2)(\mathfrak{F} + \mathfrak{F}^2) + 2\mathfrak{F}'(\mathfrak{F} + 2\mathfrak{F}\mathfrak{F}) + \mathfrak{F} + 2\mathfrak{F}^2 + 2\mathfrak{F}\mathfrak{F}) \quad \text{.........(xxx),}\end{aligned}$$

or $$\begin{aligned}&= \sigma^2_q\sigma^2_{q'} + \sigma^2_q\bar{x}^2_{q'} + \sigma^2_{q'}\bar{x}^2_q + \bar{x}^2_q\bar{x}^2_{q'} + 2\bar{x}_{q'}\mu_3(q) + 4\bar{x}_{q'}\bar{x}_q\sigma^2_q + \mu_4(q) - 3\sigma^4_q \\ &\qquad + 2\sigma^4_q + 2\bar{x}_q\mu_3(q) \quad \text{......(xxx}^{\text{bis}}\text{).}\end{aligned}$$

By aid of (xxviii$^{\text{bis}}$), (xxix$^{\text{bis}}$) and (xxx$^{\text{bis}}$) we find

$$\begin{aligned}
\mu_4'{}_{(q'-q)} &= \{(x_{q'}-x_q)^4\} \\
&= \{x^4_{q'}\} - 4\{x^3_{q'}x_q\} + 6\{x^2_{q'}x^2_q\} - 4\{x_{q'}x^3_q\} + \{x^4_q\} \\
&= \mu_4'(q') - 4\mu_3(q')\,\bar{x}_q - 12\bar{x}_q\bar{x}_{q'}\sigma^2_{q'} - 4\bar{x}^3_{q'}\bar{x}_q - 12\sigma^2_q\sigma^2_{q'} - 12\bar{x}^2_{q'}\sigma^2_q \\
&\quad - 12\bar{x}_{q'}\mu_3(q) - 4\mu_4(q) + 12\sigma^4_q + 6\sigma^2_q\sigma^2_{q'} + 6\bar{x}^2_{q'}\sigma^2_q + 6\bar{x}^2_q\sigma^2_{q'} + 6\bar{x}^2_q\bar{x}^2_{q'} \\
&\quad + 12\bar{x}_{q'}\mu_3(q) + 24\bar{x}_{q'}\bar{x}_q\sigma^2_q + 6\mu_4(q) - 6\sigma^4_q + 12\bar{x}_q\mu_3(q) - 4\bar{x}_{q'}\mu_3(q) \\
&\quad - 12\bar{x}_{q'}\bar{x}_q\sigma^2_q - 4\bar{x}_{q'}\bar{x}^3_q - 4\mu_4(q) - 12\bar{x}_q\mu_3(q) - 12\bar{x}^2_q\sigma^2_q + \mu_4'(q).
\end{aligned}$$

Hence remembering that $\mu_4' = \mu_4 + 4\mu_3\bar{x} + 6\mu_2\bar{x}^2 + \bar{x}^4$ we have, on rearranging,

$$\begin{aligned}
\mu_4'{}_{(q'-q)} &= \{(x_{q'}-x_q)^4\} = \mu_4(q') - 3\sigma^4_{q'} - (\mu_4(q) - 3\sigma^4_q) + 3(\sigma^2_q - \sigma^2_q)^2 \\
&\quad + 4(\mu_3(q') - \mu_3(q))(\bar{x}_{q'} - \bar{x}_q) + 6(\sigma^2_{q'} - \sigma^2_q)(\bar{x}_{q'} - \bar{x}_q)^2 + (\bar{x}_{q'} - \bar{x}_q)^4 \ldots\text{(xxxi)}.
\end{aligned}$$

Now

$$\begin{aligned}
\mu_{4\,(q'-q)} &= \{(x_{q'} - x_q - (\bar{x}_q - x_q))^4\} \\
&= \{(x_{q'} - x_q)^4\} - 4\{(x_{q'} - x_q)^3\}(\bar{x}_q - \bar{x}_q) + 6\{(x_{q'} - x_q)^2\}(\bar{x}_{q'} - \bar{x}_q)^2 - 3(\bar{x}_{q'} - \bar{x}_q)^4 \\
&= \{(x_{q'} - x_q)^4\} - 4[(\mu_3(q') - \mu_3(q))(\bar{x}_q - \bar{x}_q) + 3(\sigma^2_{q'} - \sigma^2_q)(\bar{x}_{q'} - \bar{x}_q)^2 + (\bar{x}_{q'} - \bar{x}_q)^4] \\
&\quad + 6[(\sigma^2_{q'} - \sigma^2_q)(\bar{x}_{q'} - \bar{x}_q)^2 + (\bar{x}_{q'} - \bar{x}_q)^4] - 3(\bar{x}_{q'} - \bar{x}_q)^4 \\
&= \{(x_{q'} - x_q)^4\} - 4(\mu_3(q') - \mu_3(q))(\bar{x}_{q'} - \bar{x}_q) - 6(\sigma^2_{q'} - \sigma^2_q)(\bar{x}_{q'} - \bar{x}_q)^2 - (\bar{x}_{q'} - \bar{x}_q)^4.
\end{aligned}$$

Hence substituting from (xxxi) we find

$$\begin{aligned}
\mu_{4\,(q'-q)} &= \mu_4(q') - 3\sigma^4_{q'} - (\mu_4(q') - 3\sigma^4_q) + 3(\sigma^2_{q'} - \sigma^2_q)^2 \\
&= \sigma^4(\mathfrak{h} - \mathfrak{h} + 3(\mathfrak{h}' - \mathfrak{h})^2) \ldots\ldots\ldots\ldots\ldots\text{(xxxii)}.
\end{aligned}$$

(5) *Interrank Intervals.*

Equations (xvii), (xxvii) and (xxxii) provide us with the chief constants* of the distribution of the interrank interval between ranks q and q', given $q' > q$ and $x_{q'} > x_q$. We may state them as follows, if $i_{q',q} = x_{q'} - x_q$:

$$\begin{aligned}
\bar{i}_{q'q} &= \sigma(\mathfrak{h} - \mathfrak{h}') \\
&= \sigma\left(\frac{1}{n-q} + \frac{1}{n-q-1} + \ldots + \frac{1}{n-q'+1}\right) \ldots\ldots\ldots\text{(xxxiii)}, \\
\sigma_i^2 &= \sigma^2(\mathfrak{h}' - \mathfrak{h}) \\
&= \sigma^2\left(\frac{1}{(n-q)^2} + \frac{1}{(n-q-1)^2} + \ldots + \frac{1}{(n-q'+1)^2}\right) \ldots\ldots\text{(xxxiv)}, \\
{}_i\beta_1\dagger &= (\mathfrak{h}' - \mathfrak{h})^2/(\mathfrak{h}' - \mathfrak{h})^3 \\
&= \frac{4\left(\dfrac{1}{(n-q)^3} + \dfrac{1}{(n-q-1)^3} + \ldots + \dfrac{1}{(n-q'+1)^3}\right)^2}{\left(\dfrac{1}{(n-q)^2} + \dfrac{1}{(n-q-1)^2} + \ldots + \dfrac{1}{(n-q'+1)^2}\right)^3} \ldots\ldots\text{(xxxv)},
\end{aligned}$$

* It would be worth investigating whether the sth semi-invariant is not the difference of the q'th and qth hyperbeta function of order $s \times \sigma^s$.

† In order to obtain the sign of $\sqrt{{}_i\beta_1}$ we must remember that

$${}_i\mu_3 = \sigma^3(\mathfrak{h}' - \mathfrak{h}) = \sigma^3\left(\frac{1}{(n-q)^3} + \frac{1}{(n-q-1)^3} + \ldots + \frac{1}{(n-q'+1)^3}\right)$$

and this is essentially *positive*.

$$
\begin{aligned}
{}_i\beta_2 &= 3 + \frac{\mathfrak{k}' - \mathfrak{k}}{(\mathfrak{k}' - \mathfrak{k})^2} \\
&= 3 + \frac{6\left(\dfrac{1}{(n-q)^4} + \dfrac{1}{(n-q-1)^4} + \dots + \dfrac{1}{(n-q'+1)^4}\right)}{\left(\dfrac{1}{(n-q)^2} + \dfrac{1}{(n-q-1)^2} + \dots + \dfrac{1}{(n-q'+1)^2}\right)^2} \quad \dots\text{(xxxvi)}.
\end{aligned}
$$

It is clear that the sums of these inverse powers can be obtained at once from Tables I and II by taking the difference of two sums.

If we take the qth interval, i.e. the interval between the qth and $(q+1)$th ranks we have

$$
\left.
\begin{aligned}
&\bar{i}_{q+1,q} = \sigma/(n-q) \\
&\sigma^2_{i_{q+1,q}} = \sigma^2/(n-q)^2, \quad \text{or} \quad \sigma_{i_{q+1,q}} = \sigma/(n-q) \\
&{}_i\beta_1 = 4, \qquad {}_i\beta_2 = 9
\end{aligned}
\right\} \quad \dots\dots\dots\text{(xxxvii)},
$$

or the distribution of the qth interval in samples of n from an exponential curve is another exponential curve of standard deviation $\sigma/(n-q)$*.

As illustrations of the above theory let us consider samples of 11 and 51 drawn from an exponential curve, and consider the constants of the distribution of the range in the two cases. Here $q = 1$, and $q' = n$, and accordingly we require $\underset{n-q}{S}\left(\dfrac{1}{(n-q)^t}\right)$, or, $\dfrac{1}{10^t} + \dfrac{1}{q^t} + \dots + \dfrac{1}{1^t}$ in the first case and in the second case $\dfrac{1}{50^t} + \dfrac{1}{49^t} + \dots + \dfrac{1}{1^t}$. These are given at once by our Tables.

Distribution of Range R in Samples of 11 *and* 51.

	$n = 11$	$n = 51$
$\bar{R}$	$= \sigma \times 2{\cdot}9289{,}6825$	$= \sigma \times 4{\cdot}4992{,}0534$
σ^2_R	$= \sigma^2 \times 1{\cdot}5497{,}6773$	$= \sigma^2 \times 1{\cdot}6251{,}3273$
σ_R	$= \sigma \times 1{\cdot}244{,}897$	$= \sigma \times 1{\cdot}274{,}807$
${}_R\beta_1$	$= \dfrac{4\,(1{\cdot}1975{,}3199)^2}{(1{\cdot}5497{,}6773)^3}$	$= \dfrac{4\,(1{\cdot}2018{,}6086)^2}{(1{\cdot}6251{,}3273)^3}$
	$= 1{\cdot}54111$	$= 1{\cdot}34618$
${}_R\beta_2$	$= 3 + \dfrac{6\,(1{\cdot}0820{,}3658)}{(1{\cdot}5497{,}6773)^2}$	$= 3 + \dfrac{6\,(1{\cdot}0823{,}2065)}{(1{\cdot}6251{,}3273)^2}$
	$= 5{\cdot}70308$	$= 5{\cdot}45884$

Both these distributions show very sensible skewness and marked leptokurtosis. They indicate a curve of Pearson's Type VI†, i.e. $y = y_0 (x-a)^{m_1}/x^{m_2}$ from $x = a$ to $x = \infty$. The two curves are the following:

$$
n = 11: \; y = \frac{N}{\sigma}\, 1{\cdot}378{,}451 \times 10^{33} \left(\frac{x}{\sigma} - 12{\cdot}294{,}788\right)^{3\cdot81602} \Big/ \left(\frac{x}{\sigma}\right)^{30\cdot05094},
$$

* Putting $q = 0$, we obtain the result for the distribution of x_1, the "first" distance. See *Biometrika*, Vol. xx^A. p. 223.

† $\kappa_2 = 2{\cdot}04948$ and $1{\cdot}58909$ for samples of 11 and 51 respectively. For Type VI, κ_2 must be > 1 and $< \infty$, and for Type V, $\kappa_2 = 1$. Hence the curves are Type VI, approximating to Type V.

the origin being at $14{\cdot}738,037\,\sigma$ to the left of the mean;

$$n = 51: \; y = \frac{N}{\sigma} 5{\cdot}502,329 \times 10^{27} \left(\frac{x}{\sigma} - 8{\cdot}422,628\right)^{6{\cdot}39545} \Big/ \left(\frac{x}{\sigma}\right)^{29{\cdot}56889},$$

the origin being at $11{\cdot}327,031\,\sigma$ to the left of the mean.

Figs. I and II indicate the characteristics of these curves for $N = 1000$.

The true forms of the Range Frequency Curves are given in Section (9) (ii) below; their ordinates are indicated by black circular dots on Figs. I and II, and the fit with Pearson curves calculated from the first four moments is remarkably close notwithstanding the difference in analytical form.

The values of the skewness (= (mean − mode)/standard deviation) are respectively ·5261 and ·4803, indicating only an 8 °/₀ reduction in skewness for nearly a five-fold increase in size of sample. Even for a sample of 100 we have

$${}_R\beta_1 = 1{\cdot}32231, \; {}_R\beta_2 = 5{\cdot}42928, \text{ giving skewness} = {\cdot}47456,$$

and for $n \to \infty$,

$${}_R\beta_1 = 1{\cdot}29857, \; {}_R\beta_2 = 5{\cdot}40000, \text{ and skewness} = {\cdot}42876.$$

Thus even with very large samples, the frequency curve of ranges remains extremely skew, and differs widely from the normal curve.

The sample of 51 is almost as far from a normal distribution as the sample of 11. Even if we consider $n \to \infty$, we have

$$\sigma_R = \sigma \times 1{\cdot}282,511, \text{ and as before } {}_R\beta_1 = 1{\cdot}29857, \; {}_R\beta_2 = 5{\cdot}40000,$$

for the constants of the limiting distribution. It will be seen that the standard deviation is finite, that the skewness and leptokurtosis are still considerable, but not very different from those of $n = 51$, but the mean of this finite curve has shifted to infinity since

$$\bar{R} = \sigma\left(\frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{1}\right) = \infty^* \text{ as } n \to \infty.$$

In this case $k_2 = {\cdot}97511$, actually corresponding to a Type IV Curve, but so close to unity that a Type V would probably fit as well.

(6) *Median and Interquartile Distances.*

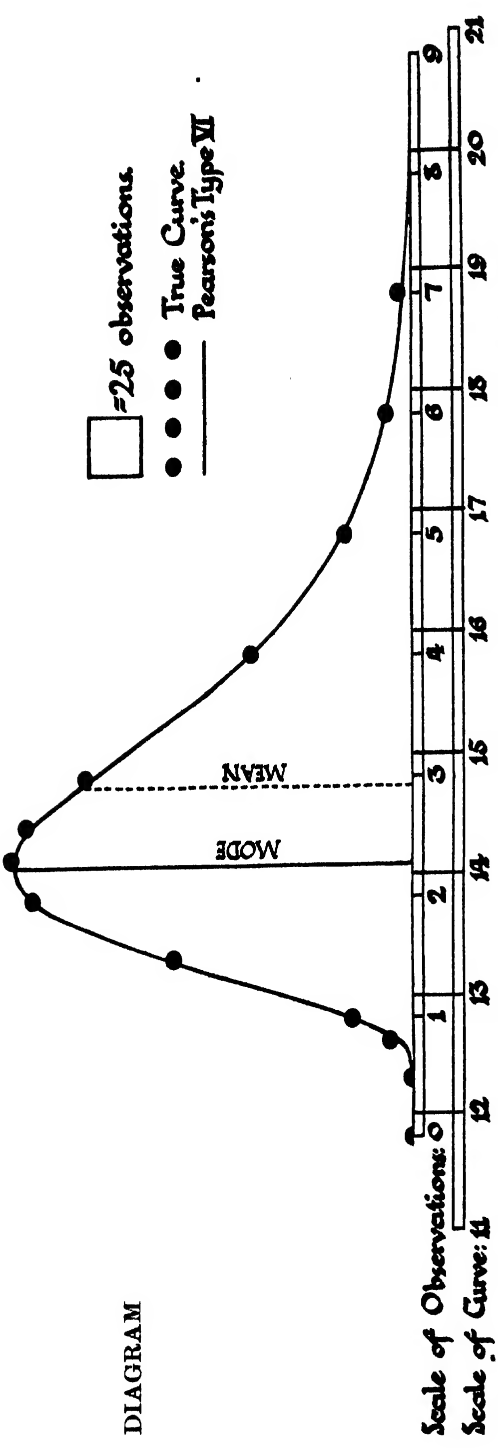
Let us now turn to the distribution of the median in samples of $4m + 3$. Here we have $q = 2m + 2$ and $n - q = 2m + 1$. We give the values for 3, 11, 25, 51, 75 and 99 for comparison. We find from (vi^bis) and our Tables:

Size of Sample	Median
3	$\sigma \times {\cdot}833,333$
11	$\sigma \times {\cdot}736,544$
23	$\sigma \times {\cdot}714,414$
51	$\sigma \times {\cdot}702,855^+$
75	$\sigma \times {\cdot}699,769$
99	$\sigma \times {\cdot}698,172$
999	$\sigma \times {\cdot}693,648$
Asymptotic value ∞	$\sigma \times {\cdot}693,147$

The last value is found from the formula on p. 205, and $= 2{\cdot}3025,8509 \log_{10} 2$.

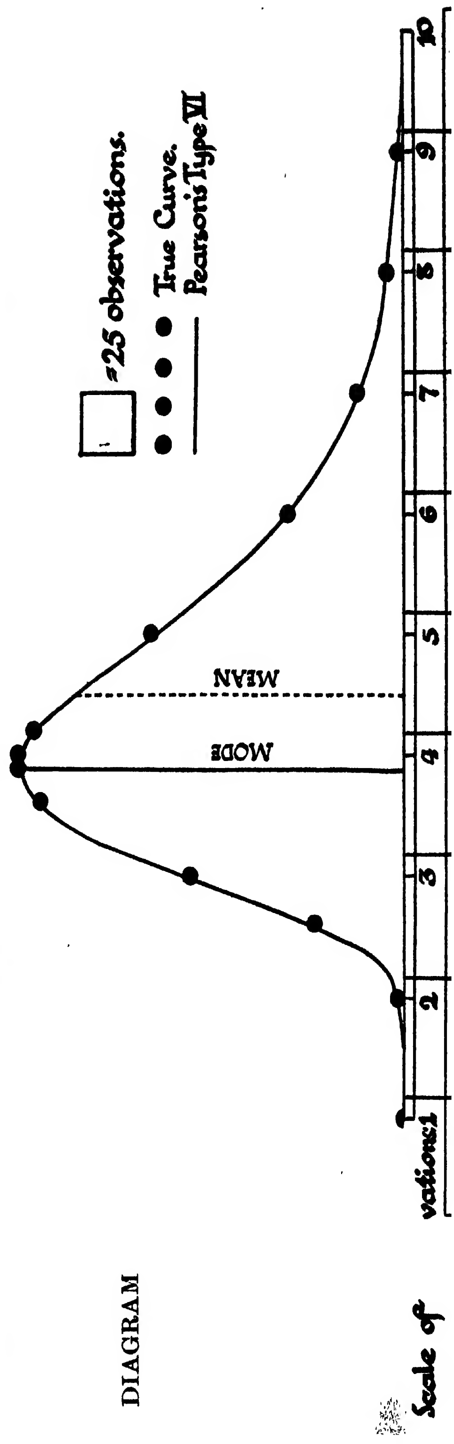
* In an infinitely great sample one individual or more from a non-limited curve like the exponential would have an infinite value, and thus the range and accordingly the mean range would be infinite.

DIAGRAM



DISTRIBUTION OF RANGE IN SAMPLES OF FIFTY-ONE FROM AN EXPONENTIAL POPULATION.

DIAGRAM



It is clear that the larger the sample the more we depart from the mean of the parent population ($\bar{x}=\sigma$) and approach nearer and nearer to the asymptotic value. It is possible, however, to determine σ and therefore the mean of the parent population from the median by dividing by the appropriate factor. It does not however follow that the median is the best single rank to find the mean from; that depends on the standard error of the rank. We will now consider the standard deviations of the median distributions.

Size of Sample	Standard Deviation of Median	Standard Deviation of Mean as found from Median
3	$\sigma \times \cdot 600,925$	$(\sigma/\sqrt{n}) \times 1\cdot 24900$
11	$\sigma \times \cdot 307,280$	$(\sigma/\sqrt{n}) \times 1\cdot 38367$
23	$\sigma \times \cdot 210,606^5$	$(\sigma/\sqrt{n}) \times 1\cdot 41379$
51	$\sigma \times \cdot 140,690$	$(\sigma/\sqrt{n}) \times 1\cdot 42949$
75	$\sigma \times \cdot 115,846$	$(\sigma/\sqrt{n}) \times 1\cdot 43370$
99	$\sigma \times \cdot 100,753$	$(\sigma/\sqrt{n}) \times 1\cdot 43586$
999	$\sigma \times \cdot 031,646^5$	$(\sigma/\sqrt{n}) \times 1\cdot 44201$
Asymptotic value ∞ *	$\sigma \times \cdot 000,000$	$(\sigma/\sqrt{n}) \times 1\cdot 44270$

These values show that relatively to the usual method of finding the mean with a standard deviation of $\sigma/\sqrt{n}$ the determination of the mean from the median is 25 % to 44 % less accurate, and the inaccuracy increases as we increase the size of the sample, although of course the *absolute* error, owing to the factor $1/\sqrt{n}$, decreases. If we ask whether the mean is better found from the range or the median, we can get a partial answer from the data on p. 219 for the distribution of range in samples of 11 and 51. We have for standard deviation of mean from range

$$=\frac{\sigma}{\sqrt{n}}\frac{\sqrt{11}\times 1\cdot 244,897}{2\cdot 9289,1825}=\frac{\sigma}{\sqrt{n}}\times 1\cdot 40966, \text{ for samples of 11,}$$

$$=\frac{\sigma}{\sqrt{n}}\frac{\sqrt{51}\times 1\cdot 274,807}{4\cdot 4992,0534}=\frac{\sigma}{\sqrt{n}}\times 2\cdot 02346, \text{ for samples of 51.}$$

Comparing these with the values $(\sigma/\sqrt{n})\times 1\cdot 38367$ and $(\sigma/\sqrt{n})\times 1\cdot 42949$, we see that the mean is better determined from the median than from the range, and the superiority of the former becomes greater as the sample increases.

We now turn to the curves of distribution of the median. These will be practically completed, if in addition to the information already provided we determine β_1 and β_2.

* This is readily obtained from the second formula on p. 205. We have

$$S\frac{1}{n^2}-S\frac{1}{(n-q)^2}=-\frac{1}{n}+\frac{1}{\frac{1}{2}(n-1)}+\text{etc.}=-\frac{1}{n}+\frac{2}{n}+\text{etc.}=\frac{1}{n}+\text{etc.}$$

$\therefore$ asymptotic value of S.D. of median $=\dfrac{\sigma}{\sqrt{n}}$, and of mean $=\dfrac{\sigma}{\sqrt{n}}\dfrac{1}{\cdot 693,147}$.

β_1 and β_2 for Distribution of Median in Samples.

Size of Sample	Value of β_1	Value of β_2
3	2·23032	6·62352
11	·75693	4·17974
23	·37720	3·58765$^-$
51	·17359	3·27026
75	·11867	3·18470
99	·08654	3·14087
999	·00896	3·01401
Asymptotic value $n \to \infty$	$9/n$	$3+14/n$
$n = \infty$	0	3

Were we dealing with the mean and not the median of a sample of n from an exponential curve, we should have the distribution curve of means with a β_1, β_2 given in terms of B_1, B_2 of the parent population by

$$\beta_1 = B_1/n = 4/n, \quad \beta_2 - 3 = (B_2 - 3)/n = 6/n.$$

Thus while both median and mean tend as n increases to be distributed normally, they do not appear to approach that state by the same route.

It may not be without interest to determine the degree of accuracy with which the mean can be found from the quartiles, confining our attention to the case of $n = 4m + 3$.

Size of Sample	Interquartile Distance	Standard Deviation of Interquartile Distance	Standard Error of Mean found from Interquartile Distance
3	$1{\cdot}500{,}000\,\sigma$	$1{\cdot}118{,}034\,\sigma$	$1{\cdot}29099\,\sigma/\sqrt{n}$
11	$1{\cdot}217{,}857\,\sigma$	$\cdot 526{,}709\,\sigma$	$1{\cdot}43440\,\sigma/\sqrt{n}$
23	$1{\cdot}156{,}219\,\sigma$	$\cdot 352{,}414\,\sigma$	$1{\cdot}46176\,\sigma/\sqrt{n}$
51	$1{\cdot}124{,}691\,\sigma$	$\cdot 232{,}346\,\sigma$	$1{\cdot}47532\,\sigma/\sqrt{n}$
75	$1{\cdot}116{,}361\,\sigma$	$\cdot 190{,}636\,\sigma$	$1{\cdot}47887\,\sigma/\sqrt{n}$
99	$1{\cdot}112{,}064\,\sigma$	$\cdot 165{,}493\,\sigma$	$1{\cdot}48070\,\sigma/\sqrt{n}$
999	$1{\cdot}101{,}152\,\sigma$	$\cdot 051{,}785\,\sigma$	$1{\cdot}48641\,\sigma/\sqrt{n}$
Asymptotic *	$(1{\cdot}098{,}612+4/3n)\,\sigma$	$\dfrac{1{\cdot}632{,}993\,\sigma}{\sqrt{n}}$	$\dfrac{1{\cdot}632{,}993\,\sigma/\sqrt{n}}{1{\cdot}098{,}612+4/3n}$
∞	$1{\cdot}098{,}612\,\sigma$	0	$1{\cdot}48641\,\sigma/\sqrt{n}$

* For the quartiles $n-q=\frac{1}{4}(3n-1)$, $n-q'=\frac{1}{4}(n-3)$ respectively.
Hence by the first formula on p. 205,

$$S\frac{1}{n-q} - S\frac{1}{n-q'} = 2{\cdot}3025{,}8509\left[\log_{10} 3 + \log_{10}\left(1-\frac{1}{3n}\right) - \log_{10}\left(1-\frac{3}{n}\right)\right] + \frac{1}{2}\frac{4}{3n-1} - \frac{1}{2}\frac{4}{n-3} - \text{etc.}$$

$$= 2{\cdot}3025{,}8509 \log_{10} 3 + \log_e\left(1-\frac{1}{3n}\right) - \log_e\left(1-\frac{3}{n}\right) + \frac{2}{3n} - \frac{2}{n} + \text{etc.}$$

$$= 1{\cdot}098{,}612 - \frac{1}{3n} + \frac{3}{n} - \frac{4}{3n} + \text{etc.}$$

$$= 1{\cdot}098{,}612 + \frac{8}{3n} - \frac{4}{3n} + \text{etc.}$$

$$= 1{\cdot}098{,}612 + \frac{4}{3n} + \text{higher inverse powers of } n.$$

It will be clear that these results are indicative of slightly less accuracy in the method of determining the mean from the interquartile distance, than in determining it from the median. But remembering what much better results are obtained for determining the median from the mean of the quartile values in the case of a normal parent population, it is worth while considering what results we obtain from the "centre" of the interquartile distance.

We first note that by (xxxiv), given *any* two ranks q and q', the (standard deviation)² of their interval i is given by

$$\sigma^2_i = \sigma^2 \left(S \frac{1}{(n-q)^2} - S \frac{1}{(n-q')^2} \right)$$

$$= \sigma^2 \left[S \frac{1}{n^2} - S \frac{1}{(n-q')^2} - \left(S \frac{1}{n^2} - S \frac{1}{(n-q)^2} \right) \right]$$

$$= \sigma^2_{q'} - \sigma^2_q, \text{ by (vii}^{\text{bis}}) \quad \ldots\ldots\ldots\ldots\text{(xxxviii)}.$$

But if $\quad c_i = \frac{1}{2}(x_q + x_{q'}), \quad \bar{c}_i = \frac{1}{2}(\bar{x}_q + \bar{x}_{q'}),$

$$c_i - \bar{c}_i = \tfrac{1}{2}(x_q - \bar{x}_q + x_{q'} - \bar{x}_{q'}),$$

then

$$\sigma^2_{c_i} = \tfrac{1}{4}(\sigma^2_q + \sigma^2_{q'} + 2\sigma_q \sigma_{q'} r_{x_q x_{q'}})$$

$$= \tfrac{1}{4}(\sigma^2_{q'} + 3\sigma^2_q),$$

or

$$\sigma_{c_i} = \tfrac{1}{2}\sqrt{\sigma^2_{q'} + 3\sigma^2_q} \quad \ldots\ldots\ldots\ldots\text{(xxxix)}.$$

Hence if $\sigma^2_q < \frac{3}{7}\sigma^2_{q'}$, σ_{c_i} will be $< \sigma_i$, and since $\bar{c}_i$ is always $> \bar{i}$ if $\bar{x}_q$ be $> \frac{1}{3}\bar{x}_{q'}$, there is some chance of better results from this method of finding the mean.

Let us see if the mean* of the parent population as found from the centre of the quartiles is more accurate than the mean as found by one of the previous methods.

Size of Sample n	Interquartile Centre c_i	Standard Deviation σ_{c_i} of c_i	Standard Error of Mean as found from c_i
3	$\sigma \times 1{\cdot}083,3333$	$\sigma \times {\cdot}650,8541$	$1{\cdot}04060\,\sigma/\sqrt{n}$
11	$\sigma \times {\cdot}910,9488$	$\sigma \times {\cdot}316,1731$	$1{\cdot}15114\,\sigma/\sqrt{n}$
23	$\sigma \times {\cdot}872,7486$	$\sigma \times {\cdot}213,6105$	$1{\cdot}17381\,\sigma/\sqrt{n}$
51	$\sigma \times {\cdot}853,2568$	$\sigma \times {\cdot}141,5987$	$1{\cdot}18513\,\sigma/\sqrt{n}$
75	$\sigma \times {\cdot}848,0664$	$\sigma \times {\cdot}116,3588$	$1{\cdot}18823\,\sigma/\sqrt{n}$
99	$\sigma \times {\cdot}845,3873$	$\sigma \times {\cdot}101,0914$	$1{\cdot}18981\,\sigma/\sqrt{n}$
999	$\sigma \times {\cdot}837,8223$	$\sigma \times {\cdot}031,6750$	$1{\cdot}19427\,\sigma/\sqrt{n}$
Asymptotic	$\sigma\left({\cdot}836,9882 + \frac{5}{6n}\right)$	$\sigma \times \frac{1}{\sqrt{n}}\left(1 + \frac{25}{48n}\right)$	$1{\cdot}19476\left(1 - \frac{{\cdot}474,800}{n}\right)\sigma/\sqrt{n}$
∞	$\sigma \times {\cdot}836,9882$	0	$1{\cdot}19476\,\sigma/\sqrt{n}$

It will be seen that the value of the mean as found from the centre of the interquartile distance is considerably better than when the mean is found from the median (p. 222), and therefore still better than when it is found from the interquartile distance itself and not from its centre (p. 223). It is not feasible with our

* The mean of the Exponential Curve $= \sigma =$ its standard deviation. Hence to find the mean of the parent population is also to find its S.D.

special definitions of the quartiles to give results for samples between 3 and 11, but if we take a sample of 5 and put $q=2$, $q'=4$, we find for the standard error of the mean as deduced from the centre of the distance between the second and fourth ranks

$$\sigma_{\text{mean}} = 1\cdot 13282\,\sigma/\sqrt{n},$$

a result which fits well into the above series. It is clear, however, that the 4 °/。 increased inaccuracy of using ranking to find the mean rapidly mounts to 13 °/。 and more, as our samples are increased in size.

(7) *Further Expansions for the Moments of Rank-Variates.*

Sufficient illustration has been provided of the relative ease with which problems in the ranking of sampled individuals from an exponential curve can be obtained from integer-argument values of the hyperbeta functions, or of the hypergamma functions as provided by our Tables I and II. But these functions have considerable interest from the standpoint of the theory of numbers. It is familiar knowledge that the sums of the inverse powers of the natural numbers can be obtained by the Euler-Maclaurin series, and the first four values are given on our p. 205. But as the moments of ranks in samples from the exponential curve can be expressed in terms of these sums, and these moments can also be obtained in series, convergent or finite, of other types, we can reach identities which may be of some interest to the pure mathematician.

In the first place, since $\alpha_x = 1 - e^{-x/\sigma}$, we have

$$\begin{aligned} x/\sigma &= -\log_e(1-\alpha_x) \\ &= \alpha_x + \frac{1}{2}\alpha_x^2 + \frac{1}{3}\alpha_x^3 + \dots + \frac{1}{s}\alpha_x^s + \dots \\ &= \overset{\infty}{\underset{s=1}{S}} (b_s\alpha_x^s), \text{ where } b_s = \frac{1}{s} \qquad \text{(xl)}. \end{aligned}$$

Again,

$$\begin{aligned} (x/\sigma)^2 &= \alpha_x^2 + \alpha_x^3 + \frac{11}{12}\alpha_x^4 + \frac{5}{6}\alpha_x^5 + \frac{137}{180}\alpha_x^6 \\ &\quad + \frac{7}{10}\alpha_x^7 + \frac{363}{560}\alpha_x^8 + \frac{761}{1{,}260}\alpha_x^9 + \frac{7{,}129}{12{,}600}\alpha_x^{10} + \frac{7{,}381}{13{,}860}\alpha_x^{11} \\ &\quad + \frac{83{,}711}{166{,}320}\alpha_x^{12} + \frac{86{,}021}{180{,}180}\alpha_x^{13} + \frac{114{,}5993}{2{,}522{,}520}\alpha_x^{14} + \dots \qquad \text{(xli)} \\ &= \overset{\infty}{\underset{s=0}{S}} c_{s+2}\alpha_x^{s+2}, \end{aligned}$$

where
$$c_{s+2} = \frac{2}{s+1}\left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{s+1}\right),$$

and
$$c_{s+2} = \frac{s+1}{s+2}c_{s+1} + \frac{}{(s+1)(s+2)} \qquad \text{.(xlii)},$$

while
$$c_1 = 0.$$

These provide easy means of computing the successive terms for checking them. It is clear that neither series converges very rapidly, and we have to employ many terms in using them.

The convergence of the terms will be increased when we introduce x and x^2 into the B-function expressions for ${}_n\bar{x}_q$ and $\mu_2'(x, q) = {}_n\sigma^2_q + {}_n\bar{x}^2_q$. We have in fact

$$\frac{{}_n\bar{x}_q}{\sigma} = \underset{s=1}{S}\, b_s \int_0^1 \alpha_x^{q-1}(1-\alpha_x)^{n-q}\alpha_x^s\, d\alpha_x \times \frac{n!}{(q-1)!\,(n-q)!}$$

$$= \underset{s=1}{S}\, b_s \frac{(q-1+s)!\,n!}{(q-1)!\,(n+s)!}$$

$$= \frac{q}{n+1} + \frac{1}{2}\frac{q(q+1)}{(n+1)(n+2)} + \frac{1}{3}\frac{q(q+1)(q+2)}{(n+1)(n+2)(n+3)}$$

$$+ \frac{1}{4}\frac{q(q+1)(q+2)(q+3)}{(n+1)(n+2)(n+3)(n+4)} + \text{etc.} \qquad \text{.........(xliii)},$$

which series must therefore by (vi^bis)

$$= \frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-q+1}.$$

Again,

$$\frac{\{{}_nx^2_q\}}{\sigma^2} = \underset{s=0}{S}\, c_{s+2}\int_0^1 \alpha_x^{q+1+s}(1-\alpha_x)^{n-q}\,d\alpha_x \times \frac{n!}{(q-1)!\,(n-q)!}$$

$$= \underset{s=0}{S}\, c_{s+2}\frac{(q+1+s)!}{(q-1)!}\,\frac{n!}{(n+s+2)!}$$

$$= c_2\frac{q(q+1)}{(n+1)(n+2)} + c_3\frac{q(q+1)(q+2)}{(n+1)(n+2)(n+3)} + c_4\frac{q(q+1)(q+2)(q+3)}{(n+1)(n+2)(n+3)(n+4)}$$

$$+ \text{etc.} \qquad \text{.........(xliv)}$$

$$= \frac{{}_n\sigma^2_q + {}_n\bar{x}^2_q}{\sigma^2} \text{ and by (vi}^{\text{bis}}\text{) and (vii}^{\text{bis}}\text{)}$$

$$= \frac{1}{n^2} + \frac{1}{(n-1)^2} + \dots + \frac{1}{(n-q+1)^2} + \left(\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-q+1}\right)^2 \qquad \text{...(xlv)},$$

and doubtless similar relations could be found for the higher moments. One further result only may be cited:

$$\frac{\{{}_nx_q \times {}_nx_{q'}\}}{\sigma^2} = \frac{q}{n+1}({}_{n+1}\bar{x}_{q'+1}) + \frac{1}{2}\frac{q(q+1)}{(n+1)(n+2)}({}_{n+2}\bar{x}_{q'+2})$$

$$+ \frac{1}{3}\frac{q(q+1)(q+2)}{(n+1)(n+2)(n+3)}({}_{n+3}\bar{x}_{q'+3}) + \text{etc.} \qquad \text{.........(xlvi)}$$

$$= \frac{{}_n\sigma^2_q + {}_n\bar{x}_q \times {}_n\bar{x}_{q'}}{\sigma^2} \text{ by (xv) if } q' \text{ be } > q.$$

Thus we have

$$\frac{q}{n+1}\left(\frac{1}{n+1} + \frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-q'+1}\right)$$

$$+ \frac{1}{2}\frac{q(q+1)}{(n+1)(n+2)}\left(\frac{1}{n+2} + \frac{1}{n+1} + \frac{1}{n} + \dots + \frac{1}{n-q'+1}\right)$$

$$+ \frac{1}{3}\frac{q(q+1)(q+2)}{(n+1)(n+2)(n+3)}\left(\frac{1}{n+3} + \frac{1}{n+2} + \frac{1}{n+1} + \dots + \frac{1}{n-q'+1}\right) + \text{etc.}$$

$$= \frac{1}{n^2} + \frac{1}{(n-1)^2} + \frac{1}{(n-2)^2} + \dots + \frac{1}{(n-q+1)^2}$$

$$+ \left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \dots + \frac{1}{2-q+1}\right)\left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \dots + \frac{1}{n-q'+1}\right).$$

Hitherto the series we have been dealing with are not finite. We will now express the infinite series in (xliii) and (xliv) by finite series.

Let the series in (xliii) be represented by

$$S = u_1 + u_2 + \ldots + u_t + \ldots,$$

where

$$\begin{aligned}
u_t &= \frac{1}{t}\,\frac{q(q+1)(q+2)\ldots(q+t-1)}{(n+1)(n+2)\ldots(n+t)} \\
&= \frac{1}{t}\,\frac{\Gamma(q+t)\,\Gamma(n+1)}{\Gamma(q)\,\Gamma(n+t+1)} = \frac{1}{t}\,\frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)} \cdot \frac{\Gamma(q+t)\,\Gamma(n-q+1)}{\Gamma(n+t+1)} \\
&= \frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)}\,\frac{1}{t}\,\mathrm{B}(q+t,\, n-q+1) \\
&= \frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n+q+1)} \int_0^1 \frac{1}{t}\, z^{q+t-1}(1-z)^{n-q}\,dz.
\end{aligned}$$

Hence
$$\begin{aligned}
S &= \frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)} \int_0^1 \sum_1^\infty \frac{1}{t}\, z^t \,.\, z^{q-1}(1-z)^{n-q}\,dz \\
&= \frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)} \int_0^1 (-\log(1-z)\, z^{q-1}(1-z)^{n-q})\,dz.
\end{aligned}$$

Take
$$(1-z) = e^{-u}, \quad dz = e^{-u}\,du,$$
and when
$$z = 0, \quad u = 0, \quad z = 1, \quad u = \infty.$$

Thus
$$\begin{aligned}
S &= \frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)} \int_0^\infty u\,(1-e^{-u})^{q-1} e^{-u\,(n-q+1)}\,du \\
&= \frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)} \int_0^\infty \sum_{s=0}^{q-1} \frac{(q-1)!\,(-1)^s}{s!\,(q-s-1)!}\, u\, e^{-u\,(n-q+s+1)}\,du.
\end{aligned}$$

Write $u' = (n-q+s+1)\,u$, and we have

$$S = \frac{n!}{(n-q)!} \sum_{s=0}^{q-1} \frac{(-1)^s}{s!\,(q-s-1)!}\, \frac{1}{(n-q+s+1)^2} \int_0^\infty u' e^{-u'}\,du',$$

where the integral is simply unity. Hence

$$\begin{aligned}
\frac{{}_n\bar{x}_q}{\sigma} &= \frac{n!}{(n-q)!\,(q-1)!} \left[\frac{1}{(n-q+1)^2} - \frac{q-1}{1!}\,\frac{1}{(n-q+2)^2} + \frac{(q-1)(q-2)}{2!}\,\frac{1}{(n-q+3)^2} \right. \\
&\qquad \left. - \frac{(q-1)(q-2)(q-3)}{3!}\,\frac{1}{(n-q+4)^2} + \ldots + (-1)^{q-1}\frac{1}{n^2} \right] \\
&= \frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \ldots + \frac{1}{n-q+1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xlvii)}.
\end{aligned}$$

Thus the series for ${}_n\bar{x}_q$ has been replaced by a finite series of q terms, which by aid of a table of the binomial coefficients and one of squares admits of easy computation. In the case of $n = 14$ and $q = 10$, we find by the computing of ten terms

$${}_{14}\bar{x}_{10} = 1{\cdot}168{,}228{,}993 \times \sigma\,*.$$

* This agrees to the whole nine decimals with the value to be found from our Table of $S(1/n)$ by simply subtracting the value for $n=4$ from that for $n=14$.

Accuracy to the fifth decimal place is only obtained by 70 to 80 terms of the series (xliv). In precisely similar manner we can obtain the kth moment coefficient of the variate of the qth rank. In this case the integral is

$$\int_0^\infty u'^k e^{-u'} du' = \Gamma(k+1),$$

and the inverse powers are the $(k+1)$th, i.e.

$$\frac{\mu_k'}{\sigma^k} = \frac{n!\,\Gamma(k+1)}{(n-q)!\,(q-1)!}\left[\frac{1}{(n-q+1)^{k+1}} - \frac{q-1}{1!}\frac{1}{(n-q+2)^{k+1}} + \frac{(q-1)(q-2)}{2!}\frac{1}{(n-q+3)^{k+1}} - \frac{(q-1)(q-2)(q-3)}{3!}\frac{1}{(n-q+4)^{k+1}} + (-1)^{q-1}\frac{1}{n^{k+1}}\right] \quad \text{.........(xlviii).}$$

In particular

$$\mu_2' = {}_n\sigma_q^2 + {}_n\bar{x}_q^2 = \frac{n!\,2!\,\sigma^2}{(n-q)!\,(q-1)!}\left[\frac{1}{(n-q+1)^3} - \frac{q-1}{1!}\frac{1}{(n-q+2)^3} + \frac{(q-1)(q-2)}{2!}\frac{1}{(n-q+3)^3} - \frac{(q-1)(q-2)(q-3)}{3!}\frac{1}{(n-q+4)^3} + \dots + (-1)^{q-1}\frac{1}{n^3}\right] \quad \text{.........(xlix).}$$

Now by (iii), reinstating the value of the standard deviation, we have

$$\frac{\mu_k'}{\sigma^k} = \frac{(-1)^k n!}{(q-1)!\,(n-q)!}\frac{d^k}{dn^k}\mathrm{B}(q, n-q+1).$$

Thus it follows that

$$\frac{d^k}{dn^k}\mathrm{B}(q, n-q+1) = (-1)^k\Gamma(k+1)\left(\frac{1}{(n-q+1)^{k+1}} - \frac{q-1}{1!}\frac{1}{(n-q+2)^{k+1}} + \frac{(q-1)(q-2)}{2!}\frac{1}{(n-q+3)^{k+1}} - \frac{(q-1)(q-2)(q-3)}{3!}\frac{1}{(n-q+4)^{k+1}} + \dots + (-1)^{q-1}\frac{1}{n^{k+1}}\right) \quad \text{......(l).}$$

I have not come across this series of the kth differential of the B-function before, but it may be known.

Since the differential coefficients of the B-function can all be expressed in terms of the hyperbeta functions, or in terms of inverse-power series, we have another long series of relations connecting the latter with a finite series of inverse powers of one higher order. Thus for example:

$$\frac{1}{n^2} + \frac{1}{(n-1)^2} + \dots + \frac{1}{(n-q+1)^2} + \left(\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-q+1}\right)^2$$
$$= \frac{n!\,2!}{(n-q)!\,(q-1)!}\left(\frac{1}{(n-q+1)^3} - \frac{q-1}{1!}\frac{1}{(n-q+2)^3} + \frac{(q-1)(q-2)}{2!}\frac{1}{(n-q+3)^3} - \frac{(q-1)(q-2)(q-3)}{3!}\frac{1}{(n-q+4)^3} + \dots + (-1)^{q-1}\frac{1}{n^3}\right) \quad \text{......(li).}$$

It is difficult to state whether these results have mathematical interest or novelty; they can be supplemented by introducing the differences of the powers of zero!

I have not succeeded in putting formula (xlvi) into a finite form.

Before developing the solution in hyperbeta functions I had applied (xliii), (xliv) and (xlvi) to a number of numerical cases with a view to obtaining a general appreciation of the nature of rank-intervals in the case of samples from an exponential curve. One such illustration is, perhaps, worth preserving, partly because we can compare results obtained in it with samples of the same size from other curves, and partly because it may have some relation to matters of some interest in sport.

(8) *Galton's Prize Ratio for the Exponential Curve.*

The distribution of runs in cricket certainly does not follow an exponential curve, but for cricket teams of about the same grade it is not widely different. The scores approximate to a J-curve with the maximum frequency at low scores and a long tail of the centuries and above.

Let the mean (or standard deviation) of the parent population be as previously σ, then

$$\begin{aligned} {}_{11}\bar{x}_{11} &= \sigma \times 3{\cdot}0198{,}7734, \\ {}_{11}\bar{x}_{1} &= \sigma \times {\cdot}0909{,}0909, \end{aligned}$$

and

$$\begin{aligned} {}_{11}\bar{x}_{2} &= \sigma \times {\cdot}1909{,}0909, \\ {}_{11}\bar{x}_{10} &= \sigma \times 2{\cdot}0198{,}7734. \end{aligned}$$

Hence

$$\begin{aligned} {}_{11}\bar{x}_{11} - {}_{11}\bar{x}_{10} &= \sigma \times 1{\cdot}0, \\ {}_{11}\bar{x}_{2} - {}_{11}\bar{x}_{1} &= \sigma \times 0{\cdot}1. \end{aligned}$$

Thus, if the average cricket score were 20, the average interval between the two lowest scores would be two runs, but between the two highest scores 20 runs. This is an illustration from a skew-curve that the general principle found for samples from normal populations still holds true, i.e. that the difference between the best individuals is much greater than that between modal individuals.

Now take

$${}_{11}\bar{x}_{9} = \sigma \times 1{\cdot}5198{,}7734,$$

and

$${}_{11}\bar{x}_{10} - {}_{11}\bar{x}_{9} = \sigma \times 0{\cdot}5.$$

Galton considered* that the ratio, R, of the value of the first to the second prize (assuming there was no third) should be as the excess of the merit of the first about the third to the excess of the merit of the second above the third, i.e.

$$R = \frac{{}_{11}\bar{x}_{11} - {}_{11}\bar{x}_{9}}{{}_{11}\bar{x}_{10} - {}_{11}\bar{x}_{9}} = \frac{1{\cdot}5}{0{\cdot}5} = 3.$$

Galton, basing his conclusions partly on experiment, partly on the normal curve, concluded that 75 °/。 and 25 °/。 of the prize money should be given to the first and second competitors. There is something, however, unique in the case of the

* *Biometrika*, Vol. I. pp. 385 *et seq.*

exponential curve: *Galton's Ratio remains constant whatever be the size of the sample**, whereas in the case of the normal curve, we only approximate to the ratio 75 °/₀ to 25 °/₀ as we indefinitely increase the size of the sample.

Proceeding a stage further we have

$$\{{}_{11}x_1^2\} = {}_{11}\sigma_1^2 + {}_{11}\bar{x}_1^2 = \cdot 0165{,}2893 \times \sigma^2,$$
$$\{{}_{11}x_2^2\} = {}_{11}\sigma_2^2 + {}_{11}\bar{x}_2^2 = \cdot 0547{,}1074 \times \sigma^2,$$
$$\{{}_{11}x_{10}^2\} = {}_{11}\sigma_{10}^2 + {}_{11}\bar{x}_{10}^2 = 4\cdot 0799{,}0447 \times \sigma^2,$$
$$\{{}_{11}x_{11}^2\} = {}_{11}\sigma_{11}^2 + {}_{11}\bar{x}_{11}^2 = 10\cdot 6776{,}9142 \times \sigma^2,$$

whence
$$ {}_{11}\sigma_1 = \cdot 0909{,}0909 \times \sigma, \quad {}_{11}\sigma_2 = \cdot 1351{,}4608 \times \sigma,$$
$$ {}_{11}\sigma_{10} = \cdot 7470{,}1552 \times \sigma, \quad {}_{11}\sigma_{11} = 1\cdot 2482{,}1163 \times \sigma.$$

Further,
$$\{{}_{11}x_1 \times {}_{11}x_2\} = \cdot 0256{,}1983 \times \sigma^2,$$
$$\{{}_{11}x_1 \times {}_{11}x_{11}\} = \cdot 2827{,}9877 \times \sigma^2.$$

These latter require an inordinate number of terms of (xlvi) to compute them to eight figure accuracy.

But
$$r_{{}_{11}x_1, {}_{11}x_2} = \frac{\{{}_{11}x_1 \times {}_{11}x_2\} - {}_{11}\bar{x}_1 \times {}_{11}\bar{x}_2}{{}_{11}\sigma_1 \times {}_{11}\sigma_2}$$
$$= \frac{\cdot 0256{,}1983 - \cdot 0909{,}0909 \times \cdot 1909{,}0909}{\cdot 0909{,}0909 \times \cdot 1351{,}4608}$$
$$= \cdot 672{,}672,$$

agreeing practically with the value

$$r_{{}_{11}x_1, {}_{11}x_2} = {}_{11}\sigma_1 / {}_{11}\sigma_2 = \frac{\cdot 0909{,}0909}{\cdot 1351{,}4608} = \cdot 672{,}673$$

found by the recent shorter method of (xvi^{bis}).

Similarly
$$r_{{}_{11}x_1, {}_{11}x_{11}} = \frac{\{{}_{11}x_1 \times {}_{11}x_{11}\} - {}_{11}\bar{x}_1 \times {}_{11}\bar{x}_{11}}{{}_{11}\sigma_1 \times {}_{11}\sigma_{11}}$$
$$= \frac{\cdot 2827{,}9877 - \cdot 0909{,}0909 \times 3\cdot 0198{,}7724}{\cdot 0909{,}0909 \times 1\cdot 2482{,}1163}$$
$$= \cdot 072{,}8315,$$

while
$${}_{11}\sigma_1 / {}_{11}\sigma_{11} = \cdot 072{,}8316,$$

only differing in the 7th decimal place.

We see that the score of the 1st rank is highly correlated with the score of the 2nd rank, but its correlation with the score of the 11th rank is relatively small. If we may suppose cricket scores to be roughly given by an exponential curve, we might interpret this by saying that the best man getting a high score has not much influence on the scores of the tail of the eleven, but that the tail of the eleven will on the average score relatively well or relatively badly as a whole, their

* It should be noted that whatever be the size of the sample the interrank intervals and their standard errors remain constant for the same interval reckoned from the nth rank. Increasing the size of the sample only pushes the series of intervals further towards the tail of the exponential, introducing additional intervals behind these.

score correlations being high. The actual explanation of the phenomena lies in the fact that if the best get a high score this has to be carried down through the middle members of the team, before space is gained for the expansion of the scores of the tail, whereas if the last but one or last but two of the tail scores well there is a wider space for the remainder, which on the average will have a higher score also. Of course the correlation of the second man's score with the first man's (i.e. the variate of the 10th with that of the 11th rank) is high also; it equals ${}_{11}\sigma_{10}/{}_{11}\sigma_{11} = \cdot 598{,}469$. In the same way ${}_{11}\sigma_9 = \cdot 555{,}0065$, and the correlation of the scores of the first and third man (i.e. of 11th and 9th rank variates) falls to ·444,641, and so on down the team till it reaches ·072,832, for the scores of the first and last man.

(9) *Further Problems on Ranks and Rank-Intervals.*

(i) To find the Curve of Frequency for the qth Rank-Variate.

The probability of the qth rank occurring in the element $d\alpha_x$ is

$$\frac{(n)!}{(q-1)!\,(n-q)!}\,\alpha_x^{q-1}\,d\alpha_x\,(1-\alpha_x)^{n-q}.$$

Hence the probability of the variate of the qth rank lying between x and $x+dx$

$$=\frac{(n)!}{(q-1)!\,(n-q)!}\,\frac{1}{\sigma}\,e^{-\frac{x}{\sigma}}\,dx\,(1-e^{-\frac{x}{\sigma}})^{q-1}\,e^{-\frac{x}{\sigma}(n-q)}.$$

Or, the frequency curve for the qth Rank-Variate is

$$y=\frac{N}{\sigma}\,\frac{(n)!}{(q-1)!\,(n-q)!}\,e^{-\frac{x}{\sigma}(n-q+1)}\,(1-e^{-\frac{x}{\sigma}})^{q-1}.$$

If $q=1$, the frequency curve is

$$y=\frac{Nn}{\sigma}\,e^{-\frac{nx}{\sigma}},$$

an exponential curve of standard deviation $\frac{\;}{n}$

If $q=n$, the frequency curve is

$$y=\frac{Nn}{\sigma}\,e^{-\frac{x}{\sigma}}(1-e^{-\frac{x}{\sigma}})^{n-1}.$$

If n be even, and $q=\frac{1}{2}n+1$, the frequency curve is

$$y=\frac{Nn}{\sigma}\,[e^{-\frac{x}{\sigma}}(1-e^{-\frac{x}{\sigma}})]^{\frac{1}{2}n}.$$

This is the distribution of the "median."

(ii) To find the (q, q') Rank-Interval.

Here the probability of the qth rank occurring in $d\alpha_x$, and the qth rank in $d\alpha_{x'}$ is

$$\frac{(n)!}{(q-1)!\,(q'-q-1)!\,(n-q')!}\,\alpha_x^{q-1}\,d\alpha_x\,(\alpha_{x'}-\alpha_x)^{q'-q-1}\,(1-\alpha_{x'})^{n-q'}\,d\alpha_{x'}.$$

* See p. 212 above.

Or, the probability of the qth and q'th variates occurring in dx and dx' respectively is

$$\frac{(n)!}{(q-1)!\,(q'-q-1)!\,(n-q')!}(1-e^{-\frac{x}{\sigma}})^{q-1}e^{-\frac{x}{\sigma}}\frac{dx}{\sigma}(e^{-\frac{x}{\sigma}}-e^{-\frac{x'}{\sigma}})^{q'-q-1}e^{-\frac{x'}{\sigma}(n-q'+1)}\frac{dx'}{\sigma}.$$

Write $x'=R_{q,q'}+x$, and integrate x between the limits 0 and ∞. Then the frequency curve of ranges, $R_{q,q'}$, in N samples of size n is

$$y=\frac{N(n)!}{(q-1)!\,(q'-q-1)!\,(n-q')!}\frac{1}{\sigma^2}(1-e^{-\frac{R_{q,q'}}{\sigma}})^{q'-q-1}e^{-\frac{R_{q,q'}}{\sigma}(n-q'+1)}$$
$$\times\int_0^\infty e^{-\frac{x}{\sigma}(n-q+1)}(1-e^{-\frac{x}{\sigma}})^{q-1}dx.$$

Take $z=e^{-\frac{x}{\sigma}}$ and

$$\int_0^\infty e^{-\frac{x}{\sigma}(n-q+1)}(1-e^{-\frac{x}{\sigma}})^{q-1}dx=\sigma\int_0^1 z^{n-q}(1-z)^{q-1}dz$$
$$=\mathrm{B}(n-q+1,q)=\frac{(n-q)!\,(q-1)!}{(n)!}.$$

Thus the frequency curve becomes

$$y=\frac{N(n-q)!}{\sigma(q'-q-1)!\,(n-q')!}(1-e^{-\frac{R_{q,q'}}{\sigma}})^{q'-q-1}e^{-\frac{R_{q,q'}}{\sigma}(n-q'+1)}.$$

If $q=1$, $q'=n$, we have the frequency curve for the entire range R:

$$y=\frac{N(n-1)}{\sigma}(1-e^{-\frac{R}{\sigma}})^{n-2}e^{-\frac{R}{\sigma}}.$$

If $q'=q+1$, we have the distribution of range of the qth interval, namely

$$\frac{N(n-q)}{\sigma}e^{-\frac{R_{q,q+1}}{\sigma}(n-q)},$$

which is an exponential curve of standard deviation $\frac{\sigma}{n-q}$ with mean $R_{q,q+1}=\frac{\sigma}{n-q}$ also.

The relation of our interrank distance curve to the rank-variate curve is seen at once if we put in the former $q=0$, when it becomes

$$\frac{N(n)!}{\sigma(q'-1)!\,(n-q')!}(1-e^{-\frac{R_{0,q'}}{\sigma}})^{q'-1}e^{-\frac{R_{0,q'}}{\sigma}(n-q'+1)},$$

which is identical with the curve for the q'th rank-variate, if we take $R_{0,q'}=x_{q'}$, i.e. take the zero-rank to be the start of the parent population.

To indicate the relation to the hyperbeta functions, we have for the sth moment coefficient about zero, writing u for $R_{q,q'}$ for brevity,

$$\mu_s(u)=\frac{N(n-q)!}{\sigma(q'-q-1)!\,(n-q')!}\int_0^\infty u^s(1-e^{-\frac{u}{\sigma}})^{q'-q-1}e^{-\frac{u}{\sigma}(n-q'+1)}du.$$

Put $$z = e^{-\frac{u}{\sigma}}, \quad u = -\sigma \log z, \quad du = -\frac{\sigma}{z}\, dz,$$

and we find $$\mu_s(u) = \frac{\sigma^s (-1)^s N (n-q)!}{(q'-q-1)!\,(n-q')!} \frac{d^s}{du^s} \mathrm{B}(q'-q, n-q'+2),$$

which connects the sth moment of $R_{q,q'}$ with the sth hyperbeta function.

The object of the previous sections has been to demonstrate that the problem of ranks and rank-intervals can be fully solved for the exponential curve. The special interest of the exponential curve lies in the fact that it describes the distribution of the intervals between random occurrences in space or time*. Accordingly if we arrange n such intervals in order of size, we can discuss many properties of random occurrences by the formulae of this article, such as the mean and variance of these intervals, and of the same quantities for the difference of intervals. It can hardly be doubted that such knowledge may be of value in determining the degree with which a given distribution of intervals in medical or social phenomena approximate to a random series. Thus a study of samples from an exponential series has a wider interest than may at first sight be attributed to it.

(10) *Ranks and Rank-Intervals for the* (β_1, β_2) *Biquadratic Curves.*

The Rectangle and the Exponential Curve are only two special cases of the curves which occur along the Biquadratic†

$$\beta_1 (8\beta_2 - 9\beta_1 - 12)(\beta_2 + 3)^2 - (10\beta_2 - 12\beta_1 - 18)^2 (4\beta_2 - 3\beta_1) = 0 \quad \text{...(lii)}$$

in the β_1, β_2 plane.

All these curves can have their x easily expressed in terms of their α_x, and so the properties of their ranked samples can be obtained.

In my notation the following curves occur:

Type VIII. $$y = y_0 \left(1 + \frac{x}{a}\right)^{-m} \quad \text{..............................(liii).}$$

This Type is found from the Rectangular Point along the upper branch of the biquadratic loop.

The total range is from $x = 0$, where the ordinate is finite, to $x = -a$, when we have an infinite (asymptotic) ordinate.

Here $$m = 2(9 + 6\beta_1 - 5\beta_2)/(6 + 3\beta_1 - 2\beta_2) \quad \text{..................(liv),}$$

and is always less than unity.

We may write our curve by change of origin in the form

$$y = y_0' \left(\frac{x'}{a}\right)^{-m} \quad \text{................................(lv),}$$

* See Whitworth, *Choice and Chance*, 5th Ed. p. 200.

† *Phil. Trans.* Vol. 216A, pp. 435 *et seq.*

with range from $x' = 0$ to a. If σ be the standard deviation,

$$a = \sigma(2-m)\sqrt{\frac{3-m}{1-m}} \quad \text{(lvi)},$$

and the mean $\bar{x}$ is given by

$$\bar{x} = a(1-m)/(2-m) = \sigma\sqrt{(1-m)(3-m)} \quad \text{(lvii)}.$$

We have at once
$$\left.\begin{aligned} x' &= a\,(a_{x'})^{\frac{1}{1-m}} \\ \text{or} \quad \frac{x'}{\sigma} &= (2-m)\sqrt{\frac{3-m}{1-m}}\,a_{x'}^{\frac{1}{1-m}} \end{aligned}\right\} \quad \text{(lviii)}.$$

Since m is < 1, $\frac{1}{1-m}$ is a positive power > 1. At $m = 0$, the Rectangle is reached, and this has been fully discussed*.

Type IX_1.
$$y = y_0\left(1 + \frac{x}{a}\right)^m \quad \text{(lix)}.$$

Passing from the Rectangle Point m is positive and less than unity until we reach the Line Point L, at which m is unity. Beyond this we have m greater than unity (Type IX_2) until m equals infinity, and we have the Exponential Curve already dealt with.

We can throw these curves together as

$$y = y_0\left(\frac{x'}{a}\right)^m \quad \text{(lx)},$$

with a range from zero ordinate at $x' = 0$ to a finite ordinate at $x' = a$.

Here
$$\bar{x} = a(m+1)/(m+2) = \sigma\sqrt{(m+1)(m+3)} \quad \text{(lxi)},$$

$$a = \sigma(m+2)\sqrt{\frac{m+3}{m+1}} \quad \text{(lxii)},$$

and
$$x' = a\,(a_{x'})^{\frac{1}{m+1}} = \sigma(m+2)\sqrt{\frac{m+3}{m+1}}\,a_{x'}^{\frac{1}{m+1}} \quad \text{(lxiii)}.$$

Clearly the power of $a_{x'}$ is a positive proper fraction.

Finally
$$m = \frac{2(5\beta_2 - 6\beta_1 - 9)}{3\beta_1 - 2\beta_2 + 6} \quad \text{(lxiv)}.$$

Type XI. This is given by
$$y = y_0\left(\frac{x}{a}\right)^{-m} \quad \text{(lxv)}.$$

Passing the Exponential Point, the curve takes the above form, where the range is from $x = a$, a constant, to ∞, and

$$m = \frac{2(5\beta_2 - 6\beta_1 - 9)}{2\beta_2 - 3\beta_1 - 6} \quad \text{(lxvi)},$$

and can take all positive values from ∞ to 5.

* *Biometrika*, Vol. XXIII. pp. 390—397.

Here
$$\left.\begin{aligned}\bar{x} &= a(m-1)/(m-2) = \sigma\sqrt{(m-1)(m-3)}\\ a &= \sigma(m-2)\sqrt{\frac{m-3}{m-1}}\end{aligned}\right\} \quad \text{............(lxvii)},$$

and
$$\alpha_x = 1 - \left(\frac{a}{x}\right)^{m-1} \quad \text{..............................(lxviii)}.$$

Hence
$$x = a\,\frac{1}{(1-\alpha_x)^{\frac{1}{m-1}}} \quad \text{..(lxix)}$$
$$= \sigma(m-2)\sqrt{\frac{m-3}{m-1}}\,\frac{1}{(1-\alpha_x)^{\frac{1}{m-1}}} \quad \text{......................(lxx)}.$$

Since m is positive and greater than unity, $\dfrac{1}{m-1}$ is a positive proper fraction*.

We will now consider these curves in order in relation to ranking.

Type VIII. Let ${}_n\bar{x}_q$ be the mean of the qth rank-variate in samples of size n, and ${}_n\sigma_q$ its standard deviation, where we may drop the prefix n if there be no danger of confusion.

(i) To find ${}_n\bar{x}_q$:

$${}_n\bar{x}_q = a\,\frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)}\int_0^1 \alpha_x^{q-1}(1-\alpha_x)^{n-q}\,x_q\,d\alpha_x$$
$$= a\,\frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)}\int_0^1 \alpha_x^{q-1+\frac{1}{1-m}}(1-\alpha_x)^{n-q}\,d\alpha_x$$
$$= a\,\frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)}\,\mathrm{B}\left(q+\frac{1}{1-m},\, n-q+1\right)$$
$$= a\,\frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)}\,\frac{\Gamma\left(q+\frac{1}{1-m}\right)\Gamma(n-q+1)}{\Gamma\left(n+\frac{1}{1-m}+1\right)}$$
$$= a\,\frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{1}{1-m}\right)}\,\frac{\Gamma\left(q+\frac{1}{1-m}\right)}{\Gamma(q)} \quad \text{......................(lxxi)}.$$

This can be found from any table of Γ-functions.

$$\mu_2'(n,q)\,{}_n\sigma_q^2 + {}_n\bar{x}_q^2 = \frac{\Gamma(n+1)}{\Gamma(q)\,\Gamma(n-q+1)}\int_0^1 \alpha_x^{q-1}(1-\alpha_x)^{n-q}\,x_q^2\,d\alpha_x$$
$$\frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{2}{1-m}\right)}\,\frac{\Gamma\left(q+\frac{2}{1-m}\right)}{\Gamma(q)} \quad \text{.........(lxxii)}.$$

* For a fuller discussion of these curves the reader is referred to the memoir already cited. It must be remembered throughout that β_2 is given in terms of β_1 by the equation to the biquadratic.

Similarly $$\mu_s'(n, q) = a^s \frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{s}{1-m}\right)} \frac{\Gamma\left(q+\frac{s}{1-m}\right)}{\Gamma(q)} \quad \text{.........(lxxiii)},$$

from which, by reduction to the mean, all the higher moments and β's can be found.

Interesting special cases can be considered. For example the mean of the variate of the first rank, i.e. $q = 1$, is

$$_n\bar{x}_1 = \frac{a}{\left(1+\frac{1}{n(1-m)}\right)\left(1+\frac{1}{(n-1)(1-m)}\right)\left(1+\frac{1}{(n-2)(1-m)}\right)\cdots\left(1+\frac{1}{1-m}\right)},$$

and the mean of the variate of the last rank

$$_n\bar{x}_n = \frac{\quad}{n(1-m)}$$

The relation of these to the general value

$$_n\bar{x}_q \quad \frac{a}{\left(1+\frac{1}{n(1-m)}\right)\left(1+\frac{1}{(n-1)(1-m)}\right)\left(1+\frac{1}{(n-2)(1-m)}\right)\cdots\left(1+\frac{1}{q(1-m)}\right)} \quad \text{......(lxxiv)}$$

is clear.

We have difference relations of the form

$$_n\bar{x}_{q-1} = \frac{1}{1+\frac{1}{(q-1)(1-m)}} \, {}_n\bar{x}_q,$$

and $$_n\bar{x}_q = \left(1+\frac{1}{(q-1)(1-m)}\right) {}_n\bar{x}_{q-1}.$$

The latter leads us to the mean interval between adjacent ranks, $q-1$th and qth:

$$_n\bar{x}_q - {}_n\bar{x}_{q-1} = \frac{1}{(q-1)(1-m)} \, {}_n\bar{x}_{q-1}$$

$$= \frac{a}{(q-1)(1-m)} \frac{1}{\left(1+\frac{1}{n(1-m)}\right)\left(1+\frac{1}{(n-1)(1-m)}\right)\cdots\left(1+\frac{1}{(q-1)(1-m)}\right)}.$$

Again

$$_n\bar{x}_{q-1} - {}_n\bar{x}_{q-2}$$

$$= \frac{a}{(q-2)(1-m)} \frac{1}{\left(1+\frac{1}{n(1-m)}\right)\left(1+\frac{1}{(n-1)(1-m)}\right)\cdots\left(1+\frac{1}{(q-2)(1-m)}\right)}.$$

Or $$\frac{_n\bar{x}_q - {}_n\bar{x}_{q-1}}{_n\bar{x}_{q-1} - {}_n\bar{x}_{q-2}} = \frac{q-2}{q-1}\left(1+\frac{1}{(q-2)(1-m)}\right) = \frac{q-2}{q-1} + \frac{1}{(q-1)(1-m)} \quad \text{......(lxxv)}.$$

It is clear from this result that the ratio of interrank distances is independent of the size of the sample, but their actual values and their position do depend on the size of the sample. Further we have

$$\frac{{}_n\bar{x}_q - {}_n\bar{x}_{q-2}}{{}_n\bar{x}_{q-1} - {}_n\bar{x}_{q-2}} = \frac{2q-3}{q-1} + \frac{1}{(q-1)(1-m)}.$$

Putting in succession $q = n$ and $q = 3$, we have

$$\frac{{}_n\bar{x}_n - {}_n\bar{x}_{n-2}}{{}_n\bar{x}_{n-1} - {}_n\bar{x}_{n-2}} = \frac{2n-3}{n-1} + \frac{1}{(n-1)(1-m)} = 2 + \frac{1}{n-1}\frac{m}{1-m} \quad \text{...(lxxvi)},$$

and
$$\frac{{}_n\bar{x}_3 - {}_n\bar{x}_2}{{}_n\bar{x}_2 - {}_n\bar{x}_1} = \frac{1}{2} + \frac{1}{2(1-m)} = \frac{1}{2}\frac{2-m}{1-m},$$

$$\frac{{}_n\bar{x}_3 - {}_n\bar{x}_1}{{}_n\bar{x}_2 - {}_n\bar{x}_1} = \frac{1}{2}\frac{2-m}{1-m} + 1 = \frac{1}{2}\frac{4-3m}{1-m},$$

and thus
$$\frac{{}_n\bar{x}_3 - {}_n\bar{x}_1}{{}_n\bar{x}_3 - {}_n\bar{x}_2} = \frac{4-3m}{2-m} = 2 - \frac{m}{2-m} \quad \text{.............................(lxxvii)}.$$

Comparing (lxxvi) with (lxxvii) we see that Galton's Ratio is a little greater than 2 at the stump ($x = a$) and somewhat less than 2 at the asymptotic end ($x = 0$). It is difficult to say which must be looked on as the prize, which the penalty end of the rank series! We see that when $m = 0$ we get the rectangle, and both ends of the series give an identical result, 2, for Galton's Ratio.

To complete the requisite formulae we may determine the value about the origin of the curve of a product moment coefficient of any order. Writing χ for $\Gamma(n+1)/[\Gamma(q)\,\Gamma(q'-q)\,\Gamma(n-q'+1)]$, we have

$$p'_{st} = \{{}_nx_q{}^s \times {}_nx_{q'}{}^t\} = a^{s+t}\chi\int_0^1 d\alpha_x \alpha_x{}^{q-1}\int_{\alpha_x}^1 d\alpha_{x'}(\alpha_{x'}-\alpha_x)^{q'-q-1}(1-\alpha_{x'})^{n-q'}\alpha_x{}^{\frac{s}{1-m}}\alpha_{x'}{}^{\frac{t}{1-m}}$$

$$= a^{s+t}\chi\int_0^1 d\alpha_x \alpha_x{}^{q-1+\frac{s}{1-m}}u(q'-q-1) \quad \text{....................(lxxviii)},$$

where
$$u(q'-q-1) = \int_{\alpha_x}^1 d\alpha_{x'}(\alpha_{x'}-\alpha_x)^{q'-q-1}(1-\alpha_{x'})^{n+q'}\alpha_{x'}{}^{\frac{t}{1-n}}$$

Now

$$\frac{du(q'-q-1)}{d\alpha_x} = -(q'-q-1)\int_{\alpha_x}^1 d\alpha_{x'}(\alpha_{x'}-\alpha_x)^{q'-q-2}(1-\alpha_{x'})^{n-q'}\alpha_{x'}{}^{\frac{t}{1-m}}.$$

Integrate (lxxviii) by parts and we have

$$p'_{st} = a^{s+t}\chi\frac{q'-q-1}{q+\dfrac{s}{1-m}}\int_0^1 d\alpha_x \alpha_x{}^{q+\frac{s}{1-m}}u(q'-q-2).$$

If we repeat this integration by parts $q'-q-2$ times we find the factor $\alpha_{x'} - \alpha_x$ will disappear, the part outside the integration always vanishing. Thus

$$p'_{st} = a^{s+t}\chi \frac{q'-q-1}{q+\frac{s}{1-m}} \frac{q'-q-2}{q+1+\frac{s}{1-m}} \cdots \frac{}{q'-2+\frac{s}{1-m}}$$

$$\times \int_0^1 d\alpha_x \alpha_x^{q'-2+\frac{s}{1-m}} \int_{\alpha_x}^1 \alpha_{x'}^{\frac{t}{1-m}} (1-\alpha_{x'})^{n-q'} d\alpha_{x'}.$$

Repeating the integration by parts once more we have

$$p'_{st} = a^{s+t}\chi \frac{\Gamma(q'-q)}{\Gamma\left(q'+\frac{s}{1-m}\right)} \Gamma\left(q+\frac{s}{1-m}\right) \int_0^1 \alpha_{x'}^{q'-1+\frac{s+t}{1-m}} (1-\alpha_{x'})^{n-q'} d\alpha_{x'}.$$

The integral is now the B-function $B\left(q'+\frac{s+t}{1-m}, n-q'+1\right)$.

Replacing the B-function by Γ-functions and inserting the value of χ, we have

$$\{{}_n x_q{}^s \times {}_n x_{q'}{}^t\} = a^{s+t} \frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{s+t}{1-m}\right)} \frac{\Gamma\left(q+\frac{s}{1-m}\right)}{\Gamma(q)} \frac{\Gamma\left(q'+\frac{s+t}{1-m}\right)}{\Gamma\left(q'+\frac{s}{1-m}\right)} \quad \text{.........(lxxix).}$$

If we put $q'=q$, and $t=0$, we obtain the value of $\{{}_n x_q{}^s\} = \mu_s'(n, q)$ agreeing with the result in (lxxiii). Expanding the Γ-functions and rearranging, we have

$$p'_{st}(n, q, q') = \{{}_n x_q{}^s \times {}_n x_{q'}{}^t\}$$

$$= \frac{a^{s+t}}{\left(1+\frac{s+t}{n(1-m)}\right)\left(1+\frac{s+t}{(n-1)(1-m)}\right)\cdots\left(1+\frac{s+t}{q'(1-m)}\right)}$$

$$\times \frac{1}{\left(1+\frac{s}{(q'-1)(1-m)}\right)\left(1+\frac{s}{(q'-2)(1-m)}\right)\cdots\left(1+\frac{s}{q(1-m)}\right)} \quad \text{... (lxxx)}$$

$$= \mu'_{s+t}(n, q') \times \mu_s'(q'-1, q)/a^s \quad \text{..(lxxxi).}$$

As a particular case, if $s=t=1$,

$$\{{}_n x_q \times {}_n x_{q'}\} = {}_{q'-1}\bar{x}_q \times \mu_2'(n, q')/a$$

$$\cdot \frac{a}{\left(1+\frac{2}{n(1-m)}\right)\left(1+\frac{2}{(n-1)(1-m)}\right)\cdots\left(1+\frac{2}{q'(1-m)}\right)}$$

$$\times \frac{a}{\left(1+\frac{1}{(q'-1)(1-m)}\right)\left(1+\frac{1}{(q'-2)(1-m)}\right)\left(1+\frac{1}{q(1-m)}\right)}$$

$$= {}_n\bar{x}_{q'}\left(\tfrac{1}{2}(1-m)\right) \times {}_{q'-1}\bar{x}_q(m) \quad \text{..(lxxxii),}$$

where ${}_u\bar{x}_v(w)$ = the mean of the vth rank in a sample of size u, from a curve with power w.

It is possible to write down at once the frequency curve for the variate of the qth rank x_q. The chance of obtaining α_x

$$=\frac{(n)!}{(q-1)!\,(n-q)!}\,\alpha_x^{q-1}(1-\alpha_x)^{n-q}d\alpha_x,$$

but $\alpha_x=\left(\frac{x}{a}\right)^{1-m}$, therefore

$$\text{Chance of obtaining } x_q=\frac{1-m}{a}\,\frac{(n)!}{(q-1)!\,(n-q)!}\left(\frac{x}{a}\right)^{q(1-m)-1}\left(1-\left(\frac{x}{a}\right)^{1-m}\right)^{n-q}dx.$$

Or, the frequency distribution for N samples of n of the variate of the qth rank is

$$y=\frac{N(1-m)}{a}\,\frac{(n)!}{(q-1)!\,(n-q)!}\left(\frac{x}{a}\right)^{q(1-m)-1}\left(1-\left(\frac{x}{a}\right)^{1-m}\right)^{n-q}.$$

This is a "transformed" Type I equation. If we multiply by x^s, and transform by writing $z=(x/a)^{1-m}$, we obtain on integration the value of μ_s' given in Formula (lxxiii). Putting $m=0$ we obtain the curve for sampling from a Rectangular Population, as given (with the mode as origin) in the first section of this paper*.

By changing $1-m$ to $1+m$, we obtain the corresponding frequency curve for Type X, and by writing $-(m-1)$ for $1-m$ (m now >1), we obtain that for Type XI.

A short proof of the curve of distribution of the rank-variate may be obtained as follows:

The curve of frequency for a rectangle is

$$y=\frac{N}{a}$$

and the element of frequency is $N\,dx/b$.

Tranform this curve, the origin being at one terminal, by taking

$$\frac{x}{a}=\left(\frac{a}{x'}\right)^{m-1},\quad\text{or}\quad\frac{dx}{a}=\left(\frac{x'}{a}\right)^{-m}(1-m)\,dx'.$$

The transformed curve is now

$$y=\frac{N(1-m)}{a}\left(\frac{a}{x'}\right)^{m},$$

the form of our parent population. Since x' is fixed by x, the rank of x' will be the same as that of x, but the frequency between x and $x+\delta x$ of the qth rank is†

$$=\frac{N}{a}\,\frac{(n)!}{(q-1)!\,(n-q)!}\left(\frac{x}{a}\right)^{q-1}\left(1-\frac{x}{a}\right)^{n-q}dx,$$

or transforming to x',

$$=\frac{N}{a}\,\frac{(n)!}{(q-1)!\,(n-q)!}\left(\frac{x'}{a}\right)^{q(1-m)-1}\left(1-\left(\frac{x'}{a}\right)^{1-m}\right)^{n-q}dx',$$

which gives us the same curve of frequency as before.

* See *Biometrika*, Vol. XXIII. p. 391, Formula (xxii).

† By an oversight in *Biometrika*, Vol. XXIII. p. 391, Formula (xxxii), $(n-1)!$ is printed in the numerator for $n!$

Type IX. $$y = y_0 \left(\frac{x}{a}\right)^m$$

and $$x = a\,(\alpha_x)^{\frac{1}{m+1}}.$$

Here we have only to replace $\frac{1}{1-m}$ by $\frac{1}{m+1}$ in (lxxi), (lxxii) and (lxxix)—with $s = t = 1$—to obtain our results. We have

$$_n\bar{x}_q = a\,\frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{1}{m+1}\right)}\,\frac{\Gamma\left(q+\frac{1}{m+1}\right)}{\Gamma(q)} \quad\ldots\ldots\ldots\ldots\ldots\ldots\text{(lxxxiii)},$$

$$_n\sigma_q^2 + {}_n\bar{x}_q^2 = a^2\,\frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{2}{m+1}\right)}\,\frac{\Gamma\left(q+\frac{2}{m+1}\right)}{\Gamma(q)} \quad\ldots\ldots\ldots\ldots\ldots\ldots\text{(lxxxiv)},$$

$$\{_n x_q \times {}_n x_{q'}\} = a^2\,\frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{2}{m+1}\right)}\,\frac{\Gamma\left(q'+\frac{2}{m+1}\right)}{\Gamma\left(q'+\frac{1}{m+1}\right)}\,\frac{\Gamma\left(q+\frac{1}{m+1}\right)}{\Gamma(q)} \quad\text{(lxxxv)}.$$

These may be replaced by

$$_n\bar{x}_q = \frac{}{\left(1+\frac{1}{n(m+1)}\right)\left(1+\frac{1}{(n-1)(m+1)}\right)\cdots\left(1+\frac{1}{q(m+1)}\right)}$$
$$\ldots\ldots\ldots\text{(lxxxvi)},$$

$$\mu_2' = {}_n\sigma_q^2 + {}_n\bar{x}_q^2 = \frac{a^2}{\left(1+\frac{2}{n(m+1)}\right)\left(1+\frac{2}{(n-1)(m+1)}\right)\cdots\left(1+\frac{2}{q(m+1)}\right)}$$
$$\ldots\ldots\ldots\text{(lxxxvii)},$$

$$\{_n x_q \times {}_n x_{q'}\} = {}_{q'-1}\bar{x}_q \times ({}_n\sigma_{q'}^2 + {}_n\bar{x}_{q'}^2) \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(lxxxviii)}$$

$$\frac{a}{\left(1+\frac{1}{(q'-1)(m+1)}\right)\left(1+\frac{1}{(q'-2)(m+1)}\right)\cdots\left(1+\frac{1}{q(m+1)}\right)}$$
$$\times\frac{a}{\left(1+\frac{2}{n(m+1)}\right)\left(1+\frac{2}{(n-1)(m+1)}\right)\cdots\left(1+\frac{2}{q'(m+1)}\right)}$$
$$\ldots\ldots\ldots\text{(lxxxix)}.$$

It will be noted that if $_n x_q(m)$ denote the qth rank in a sample of n from a curve with power m, then

$$_n\sigma_q^2(m) + {}_n\bar{x}_q^2(m) = a\chi\,{}_n\bar{x}_q\left(\tfrac{1}{2}(m-1)\right),$$

and $$\{_n x_q(m) \times {}_n x_{q'}(m)\} = {}_{q'-1}\bar{x}_q(m) \times {}_n\bar{x}_q\left(\tfrac{1}{2}(m-1)\right)\ldots\ldots\ldots\ldots\ldots\text{(xc)}.$$

Precisely as in the case of Type VIII we have

$$\{_n x_q^s \times _n x_{q'}^t\} = a^{s+t} \frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{s+t}{m+1}\right)} \frac{\Gamma\left(q+\frac{s}{m+1}\right)}{\Gamma(q)} \frac{\Gamma\left(q'+\frac{s+t}{m+1}\right)}{\Gamma\left(q'+\frac{s}{m+1}\right)}$$

$$= \frac{a_{s+t}}{\left(1+\frac{s+t}{n(m+1)}\right)\left(1+\frac{s+t}{(n-1)(m+1)}\right)\cdots\left(1+\frac{s+t}{q'(m+1)}\right)}$$

$$\left(1+\frac{s}{(q'-1)(m+1)}\right)\left(1+\frac{s}{(q'-2)(m+1)}\right)\cdots\left(1+\frac{s}{q(m+1)}\right) \quad \text{.........(xci)}$$

$$= \mu'_{s+t}(n, q') \times \mu_s'(q'-1, q)/a^s \quad \text{.........(xcii)},$$

and as particular cases we can deduce the three results already given by putting $q' = q$ and (i) $s=1$, $t=0$, (ii) $s=2$, $t=0$, and (iii) $s=t=1$.

All the results for Type VIII hold for Type IX, if we replace the $1-m$ in the former case by $m+1$ in the latter. We must remember that m may now range from zero to infinity.

Type XI. Here the relationship between x and a_x changes its form and our fundamental equations also:

$$_n\bar{x}_q = a \frac{\Gamma(n+1)}{\Gamma(q)\Gamma(n-q+1)} \int_0^1 a_x^{q-1}(1-a_x)^{n-q-\frac{1}{m-1}} da_x$$

$$= a \frac{\Gamma(n+1)}{\Gamma(q)\Gamma(n-q+1)} \frac{\Gamma(q)\Gamma\left(n-q+1-\frac{1}{m-1}\right)}{\Gamma\left(n+1-\frac{1}{m-1}\right)}$$

$$= a \frac{\Gamma(n+1)}{\Gamma(n-q+1)} \frac{\Gamma\left(n-q+1-\frac{1}{m-1}\right)}{\Gamma\left(n+1-\frac{1}{m-1}\right)} \quad \text{.........(xciii)}$$

$$= \frac{a}{\left(1-\frac{1}{n(m-1)}\right)\left(1-\frac{1}{(n-1)(m-1)}\right)\left(1-\frac{1}{(n-2)(m-1)}\right)\cdots\left(1-\frac{1}{(n-q+1)(m-1)}\right)} \quad \text{.........(xciv)}.$$

Similarly

$$\mu_2'(n, q)\, _n\sigma_q^2 + _n\bar{x}_q^2$$

$$= a^2 \frac{\Gamma(n+1)}{\Gamma(n-q+1)} \frac{\Gamma\left(n-q+1-\frac{2}{m-1}\right)}{\Gamma\left(n+1-\frac{2}{m-1}\right)}$$

$$= \frac{a^2}{\left(1-\frac{2}{n(m-1)}\right)\left(1-\frac{2}{(n-1)(m-1)}\right)\left(1-\frac{2}{(n-2)(m-1)}\right)\cdots\left(1-\frac{2}{(n-q+1)(m-1)}\right)} \quad \text{.........(xcv)}.$$

In the same way it is perfectly easy to write down the expression for all the higher moment coefficients about the start of the curve at $x = a$. For we have

$$\mu_s'(n, q) = \{{}_n x_q^s\}$$

$$= a^s \frac{\Gamma(n+1)}{\Gamma\left(n+1-\frac{s}{m-1}\right)} \frac{\Gamma\left(n-q+1-\frac{s}{m-1}\right)}{\Gamma(n-q+1)} \quad \text{.....................(xcvi)}$$

$$= a^s \frac{1}{\left(1-\frac{s}{n(m-1)}\right)\left(1-\frac{s}{(n-1)(m-1)}\right)\cdots\left(1-\frac{s}{(n-q+1)(m-1)}\right)}$$

.........(xcvii).

A word of warning is here requisite. n is generally a fairly large number and m lies between 5 and infinity. We have, however, the factor $n-q+1$, the lowest value of which is unity, i.e. when q is the nth rank. Supposing m to approach the limit 5, $m-1 \to 4$, and accordingly with $s = 4$ the fourth moment coefficient would approach infinity, and higher moments might become negative.

As before

$${}_n\bar{x}_q - {}_n\bar{x}_{q-1} = \frac{a}{m-1} \frac{\Gamma(n+1)\,\Gamma\left(n-q+1-\frac{1}{m-1}\right)}{\Gamma(n-q+2)\,\Gamma\left(n+1-\frac{1}{m-1}\right)} \quad \text{......(xcviii).}$$

If we put $q = 2$, and $q = n$ in succession, we have

$${}_n\bar{x}_2 - {}_n\bar{x}_1 = \frac{a}{m-1} \frac{}{\left(n-\frac{1}{m-1}\right)\left(n-1-\frac{1}{m-1}\right)},$$

and

$${}_n x_n - {}_n x_{n-1} = \frac{a}{m-1} \frac{n!}{\left(n-\frac{1}{m-1}\right)\left(n-1-\frac{1}{m-1}\right)\cdots\left(1-\frac{1}{m-1}\right)}$$

$$= \frac{a}{m-1} \frac{1}{\left(1-\frac{1}{n(m-1)}\right)\left(1-\frac{1}{(n-1)(m-1)}\right)\cdots\left(1-\frac{1}{m-1}\right)}.$$

Hence

$$\frac{{}_n\bar{x}_2 - {}_n\bar{x}_1}{{}_n\bar{x}_n - {}_n\bar{x}_{n-1}} = \frac{1}{n-1}\left(1-\frac{1}{(n-2)(m-1)}\right)\left(1-\frac{1}{(n-3)(m-1)}\right)\cdots\left(1-\frac{1}{m-1}\right)$$

.........(xcix).

Hence with a curve of this type, since $m > 1$ and all the factors in the numerator must be less than unity, we see that the interval between the two highest individuals is more than $(n-1)$ times the interval between the two lowest individuals. For example, cricket scores follow a distribution not unlike that of the present curve, and we might accordingly anticipate that the average difference in score between the two best scorers in a sample of 11 would be upwards of 10 times the average difference in score between the two worse scorers.

It remains to find the mean product moments of the variates of two ranks q and q'. Using χ as before for

$$\Gamma(n+1)/\Gamma(q)\,\Gamma(q'-q)\,\Gamma(n-q'+1),$$

we have $$p'_{s,t} = \{{}_nx_q{}^s \times {}_nx_{q'}{}^t\} = a^{s+t}\chi\int_0^1 d\alpha_{x'}\,(1-\alpha_{x'})^{n-q'-\frac{t}{m-1}}\,u,$$

where $$u_{q'-q-1} = \int_0^{\alpha_{x'}} d\alpha_x\,\alpha_x^{q-1}\,(1-\alpha_x)^{-\frac{s}{m-1}}\,(\alpha_{x'}-\alpha_x)^{q'-q-1}.$$

Integrating by parts as on previous occasions, we find

$$p'_{s,t} = a^{s+t}\chi\frac{(q'-q-1)(q'-q-2)\ldots\ldots\ldots 1}{\left(n-q'+1-\frac{t}{m-1}\right)\left(n-q'+2-\frac{t}{m-1}\right)\ldots\left(n-q-1-\frac{t}{m-1}\right)}$$
$$\times\int_0^1 d\alpha_{x'}\,(1-\alpha_{x'})^{n-q-1-\frac{t}{m-1}}\int_0^{\alpha_{x'}} d\alpha_x\,\alpha_x^{q-1}\,(1-\alpha_x)^{-\frac{s}{m-1}}.$$

Integrate by parts once more and we find

$$p'_{s,t} = a^{s+t}\chi\frac{\Gamma(q'-q)\,\Gamma\left(n-q'+1-\frac{t}{m-1}\right)}{\Gamma\left(n-q+1-\frac{t}{m-1}\right)}\int_0^1 d\alpha_{x'}\,\alpha_{x'}^{q-1}\,(1-\alpha_{x'})^{n-q-\frac{s+t}{m-1}}$$

$$= a^{s+t}\frac{\Gamma(n+1)\,\Gamma(q'-q)}{\Gamma(q)\,\Gamma(q'-q)\,\Gamma(n-q'+1)}\,\frac{\Gamma\left(n-q'+1-\frac{t}{m-1}\right)}{\Gamma\left(n-q+1-\frac{t}{m-1}\right)}\,\frac{\Gamma(q)\,\Gamma\left(n-q+1-\frac{s+t}{m-1}\right)}{\Gamma\left(n+1-\frac{s+t}{m-1}\right)}$$

$$= a^{s+t}\frac{\Gamma(n+1)}{\Gamma\left(n+1-\frac{s+t}{m-1}\right)}\,\frac{\Gamma\left(n-q'+1-\frac{t}{m-1}\right)}{\Gamma(n-q'+1)}\,\frac{\Gamma\left(n-q+1-\frac{s+t}{m-1}\right)}{\Gamma\left(n-q+1-\frac{t}{m-1}\right)}\quad\ldots\ldots\ldots(\text{c})$$

$$= \frac{a^t}{\left(1-\frac{t}{(n-q)(m-1)}\right)\left(1-\frac{t}{(n-q-1)(m-1)}\right)\ldots\left(1-\frac{t}{(n-q'+1)(m-1)}\right)}$$
$$\times\frac{a^s}{\left(1-\frac{s+t}{n(m-1)}\right)\left(1-\frac{s+t}{(n-1)(m-1)}\right)\ldots\left(1-\frac{s+t}{(n-q+1)(m-1)}\right)}\quad\ldots\ldots(\text{ci}).$$

Here m can take all values from ∞ down to 5*.

(11) *Numerical Illustrations of the Biquadratic Curves.*

I propose to illustrate the formulae of the preceding sections on a variety of parent populations taken right round the biquadratic.

* See *Phil. Trans.* Vol. 216A, pp. 444—5.

Illustration I. Type VIII.

$$\text{Curve (i)}: \ y = \frac{N}{5a}\left(\frac{a}{x}\right)^{\cdot 8}.$$

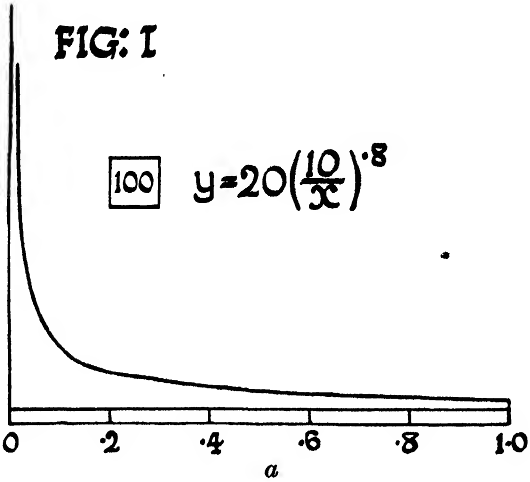
$$\text{Range } x = 0 \text{ to } x = a.$$

$$\beta_1 = 2{\cdot}75, \quad \beta_2 = 4{\cdot}714{,}2857.$$

$$\sigma = a \times {\cdot}2512{,}5945, \quad a = \sigma \times 3{\cdot}9799{,}4975.$$

The form of the curve is shown in the accompanying Fig. 1. The ranks are

FIG: I



reckoned from the asymptotic at the origin towards the stump. We suppose here, as in later curves, samples of 11 to be taken.

$$\text{Galton Ratio:} \qquad \text{at stump end} = \frac{{}_{11}\bar{x}_{11} - {}_{11}\bar{x}_9}{{}_{11}\bar{x}_{10} - {}_{11}\bar{x}_9} = 2{\cdot}4,$$

$$\text{,, \quad ,, \quad at asymptotic start} = \frac{{}_{11}\bar{x}_3 - {}_{11}\bar{x}_1}{{}_{11}\bar{x}_2 - {}_{11}\bar{x}_1} = \frac{4}{3},$$

therefore in such a case the prizes should be as 12 to 5, and the penalties as 4 to 3.

We use in the first place Equation (lxxiii) and find for $q = 1, 2, 10, 11$ and $s = 1$:

$$\frac{{}_{11}\bar{x}_1}{a} = \frac{\Gamma(12)\,\Gamma(6)}{\Gamma(16)\,\Gamma(1)} = {\cdot}0002{,}2893{,}77, \quad \frac{{}_{11}\bar{x}_2}{a} = {\cdot}0013{,}7362{,}64,$$

$$\frac{{}_{11}\bar{x}_{10}}{a} = {\cdot}4583{,}3333{,}33, \qquad \text{and} \qquad \frac{{}_{11}\bar{x}_{11}}{a} = {\cdot}6875{,}0000{,}00.$$

Dropping prefixes for brevity, we have

$$\bar{x}_2 - \bar{x}_1 = {\cdot}0011{,}4468{,}87, \quad \bar{x}_{11} - \bar{x}_{10} = {\cdot}2291{,}6666{,}67,$$

and thus
$$\frac{\bar{x}_{11} - \bar{x}_{10}}{\bar{x}_2 - \bar{x}_1} = 200{\cdot}20.$$

It is such ratios of the intervals between the best men and between two mediocre individuals which disconcert mediocrity, and make it believe that the high rankers are not the extremes of their own population, but exceptionalities belonging to a

different category, which they label "genius." It is interesting to note that the high order of this ratio is not confined to normal populations with centralised mediocrity as in the measurement of intelligence, but extends to cases like the present with mediocrity at one end of the scale as in the distribution of cricket scores, or of incomes.

Passing to the mean products of the second order, we must use (lxxiii) and (lxxix), and these give, treating a as unit,

$$\{x_1^2\} = \cdot 0000{,}0283{,}51, \quad \{x_1 x_2\} = \cdot 0000{,}0519{,}78, \quad \{x_2^2\} = \cdot 0000{,}3118{,}66,$$
$$\{x_{10}^2\} = \cdot 2619{,}0476{,}19, \quad \{x_{10} x_{11}\} = \cdot 3492{,}0634{,}92, \quad \{x_{11}^2\} = \cdot 5238{,}0952{,}38,$$
$$\{x_1 x_{10}\} = \cdot 0001{,}3082{,}16, \quad \{x_1 x_{11}\} = \cdot 0001{,}7442{,}87,$$
$$\{x_2 x_{10}\} = \cdot 0007{,}8492{,}94, \quad \{x_2 x_{11}\} = \cdot 0010{,}4657{,}25.$$

From these we deduce, restoring a:

$$\sigma^2_{x_1} = \{x_1^2\} - \bar{x}_1^2 = a^2 \times \cdot 0000{,}0278{,}27, \text{ or } \sigma_{x_1} = a \times \cdot 0016{,}6815{,}09.$$

Similarly
$$\sigma^2_{x_2} = a^2 \times \cdot 0000{,}2929{,}98, \text{ or } \sigma_{x_2} = a \times \cdot 0054{,}1292{,}01,$$
$$\sigma^2_{x_{10}} = a^2 \times \cdot 0518{,}3531{,}91, \quad \sigma_{x_{10}} = a \times \cdot 2276{,}7371{,}19,$$
$$\sigma^2_{x_{11}} = a^2 \times \cdot 0511{,}5327{,}38, \quad \sigma_{x_{11}} = a \times \cdot 2261{,}7089{,}51.$$

The *absolute* variability increases from the first rank to the tenth, and drops for the final rank. Thus at the sixth and ninth ranks we have

$$\sigma_{x_6} = \cdot 070{,}1062{,}09 \text{ and } \sigma_{x_9} = \cdot 1929{,}9154{,}70.$$

I can only explain this drop at the eleventh rank as due to the abrupt termination of the curve.

If we take *relative* variabilities as given by the coefficient of variation

$$V_t = 100\sigma_t/\bar{x}_t,$$

we have
$$V_1 = 728{\cdot}645, \quad V_2 = 394{\cdot}061, \quad V_6 = 124{\cdot}818,$$
$$V_9 = 65{\cdot}500, \quad V_{10} = 49{\cdot}674, \quad V_{11} = 32{\cdot}898,$$

or there is a continuous decrease in relative variation. The variation in the lowest rank is extremely great owing to the small amount of the character the first individual possesses, while that of the highest individual is 20 to 30 times less. Of course all the variations are high, as we are dealing only with a very small sample.

We will next consider the mean and variation of the first and last interrank intervals.

$$\sigma^2_{x_2 - x_1} = \{x_2^2\} - 2\{x_2 x_1\} + \{x_1^2\} - (\bar{x}_2 - \bar{x}_1)^2 = a^2 \times \cdot 0000{,}2231{,}58.$$

Thus
$$\sigma_{x_2 - x_1} = a \times \cdot 0047{,}2396,$$

while
$$\bar{x}_2 - \bar{x}_1 = a \times \cdot 0011{,}4469.$$

Similarly
$$\sigma^2_{x_{11} - x_{10}} = a^2 \times \cdot 0347{,}8422{,}60 \text{ and } \sigma_{x_{11} - x_{10}} = a \times \cdot 1865{,}0530,$$

while
$$\bar{x}_{11} - \bar{x}_{10} = a \times \cdot 2291{,}6667.$$

It is thus fairly easy to obtain the mean and standard deviation of any rank interval. We note that while the absolute variation of the highest interval is far greater than that of the lowest, the relative variation is much less.

We may now determine various correlations. We have, generally,

$$r_{x_q x_{q'}} = \frac{\{x_q x_{q'}\} - \{x_q\}\{x_{q'}\}}{\sigma_{x_q}\sigma_{x_{q'}}},$$

and by using the mean products we obtain for $n = 11$

$$r_{x_1 x_2} = \cdot 540{,}810{,}27, \quad r_{x_1 x_3} = \cdot 365{,}573{,}22, \quad r_{x_2 x_3} = \cdot 675{,}973{,}14,$$

and again

$$r_{x_1 x_{10}} = \cdot 068{,}173{,}14, \quad r_{x_1 x_{11}} = \cdot 045{,}148{,}76, \quad r_{x_{10} x_{11}} = \cdot 662{,}268{,}95.$$

These results indicate that the scores of the higher ranks are more closely correlated than those of the lower, but the farther the ranks correlated are apart the lower their correlation, this decreasing rapidly with the distance between them.

Further it will be found that the following relations hold:

$$r_{x_1 x_3} = r_{x_1 x_2} r_{x_2 x_3}, \quad r_{x_1 x_{11}} = r_{x_1 x_{10}} r_{x_{10} x_{11}}.$$

Seeing this in the numerical results, it drew my attention to the fact that if $q < q' < q''$, we must always have

$$r_{x_q x_{q''}} = r_{x_q x_{q'}} r_{x_{q'} x_{q''}}$$

in any system of ranks from any parent population. It is simply the statement that if you fix the variate of the q'th ranker, the partial correlation coefficient $r_{x_q x_{q''} . x_{q'}}$ must be zero; for no variation of the position of the q''th ranker can affect the qth ranker, if an intervening ranker occupies a fixed position.

We can prove this result generally for the present curve types from Formulae (lxxiii) and (lxxix) or their equivalent forms, which may be written

$$\{x_q^2\} = \mu_2' = a^2 \frac{\Gamma(n+1)}{\Gamma(n+1+2\lambda)} \frac{\Gamma(q+2\lambda)}{\Gamma(q)}, \quad \bar{x}_q = a \frac{\Gamma(n+1)}{\Gamma(n+1+\lambda)} \frac{\Gamma(q+\lambda)}{\Gamma(q)},$$

$$\{x_q x_{q'}\} = a^2 \frac{\Gamma(n+1)}{\Gamma(n+1+2\lambda)} \frac{\Gamma(q+\lambda)}{\Gamma(q)} \frac{\Gamma(q'+2\lambda)}{\Gamma(q'+\lambda)}.$$

Hence

$$\begin{aligned} r_{x_q x_{q'}} &= \frac{1}{\sigma_{x_q}\sigma_{x_{q'}}} \frac{\Gamma(n+1)\Gamma(q+\lambda)}{\Gamma(n+1+2\lambda)\Gamma(q)} \left(\frac{\Gamma(q'+2\lambda)}{\Gamma(q'+\lambda)} - \frac{\Gamma(n+1+2\lambda)\Gamma(n+1)\Gamma(q'+\lambda)}{\Gamma(n+1+\lambda)^2\Gamma(q')}\right) \\ &= \frac{1}{\sigma_{x_q}\sigma_{x_{q'}}} \frac{\Gamma(q+\lambda)}{\Gamma(q)} \frac{\Gamma(q')}{\Gamma(q'+\lambda)} \left(\frac{\Gamma(n+1)}{\Gamma(n+1+2\lambda)} \frac{\Gamma(q'+2\lambda)}{\Gamma(q')} - \left\{\frac{\Gamma(n+1)\Gamma(q'+\lambda)}{\Gamma(n+1+\lambda)\Gamma(q')}\right\}^2\right) \\ &= \frac{1}{\sigma_{x_q}\sigma_{x_{q'}}} \frac{\Gamma(q+\lambda)}{\Gamma(q)} \frac{\Gamma(q')}{\Gamma(q'+\lambda)} \left(\{x_{q'}^2\} - \{x_{q'}\}^2\right) \\ &= \frac{\sigma_{x_{q'}}}{\sigma_{x_q}} \frac{\Gamma(q+\lambda)}{\Gamma(q)} \frac{\Gamma(q')}{\Gamma(q'+\lambda)} \quad \text{..............(cii).} \end{aligned}$$

Similarly

$$r_{x_{q'} x_{q''}} = \frac{\sigma_{x_{q''}}}{\sigma_{x_{q'}}} \frac{\Gamma(q'+\lambda)}{\Gamma(q')} \frac{\Gamma(q'')}{\Gamma(q''+\lambda)},$$

and therefore

$$r_{x_q x_{q'}} r_{x_{q'} x_{q''}} = \frac{\sigma_{x_{q'}}}{\sigma_{x_q}} \frac{\Gamma(q+\lambda)}{\Gamma(q)} \frac{\Gamma(q'')}{\Gamma(q''+\lambda)} = r_{x_q x_{q''}} \quad \text{............(ciii);}$$

thus the theorem is proved generally.

If we put $\lambda = 0$, we obtain the exponential curve, and remembering the change in the sense of the axis of that curve, we obtain the result on our p. 214 (xvi^{bis}), where $\sigma_{x_{q'}}$ is now less than σ_{x_q}, the order of ranking being reversed.

As corollaries we may note that for adjacent ranks

$$r_{q,q+1} = \frac{\sigma_{q+1}}{\sigma_q}\frac{q}{q+\lambda} \quad \text{..(civ)},$$

$$\begin{aligned} r_{qq'} &= r_{q,q+1}\, r_{q,q+2}\, r_{q,q+3} \cdots r_{q'-1,q'} \\ &= \frac{q}{q+\lambda}\,\frac{q+1}{q+\lambda+1}\cdots\frac{q'-1}{q'-1+\lambda}\,\frac{\sigma_{q'}}{\sigma_q} \quad \text{.........................(cv)} \\ &= \frac{\Gamma(q')\,\Gamma(q+\lambda)}{\Gamma(q)\,\Gamma(q'+\lambda)}\,\frac{\sigma_{q'}}{\sigma_q} \text{ as before.} \end{aligned}$$

For any two rank-intervals

$$r_{q'-q,\,q'''-q''} = \frac{\left(\frac{\cdots}{\Gamma(q')} - \frac{\cdots'+\lambda)}{\Gamma(q)}\right)\left(\sigma^2_{q'''}\frac{\Gamma(q''')}{\Gamma(q'''+\lambda)} - \sigma^2_{q''}\frac{\Gamma(q'')}{\Gamma(q''+\lambda)}\right)}{\sigma_{q'-q}\times\sigma_{q'''-q''}} \quad \text{..........(cvi)},$$

which presents some simplification in computing the correlation of interrank intervals ($q < q' < q'' < q'''$).

Returning to the special example illustrated numerically on p. 245. I wrote down the regression lines of x_{10} on x_{11} and of x_1 on x_{11}, and giving various values to x_{11}, I invariably found that the percentage increase of x_{10} on $\bar{x}_{10}$ or x_1 on $\bar{x}_1$ was precisely the percentage by which I had increased x_{11} on $\bar{x}_{11}$. In other words I was forced to the conclusion that the general regression line of x_q on $x_{q'}$ must be

$$\tilde{x}_q = \frac{\bar{x}_q}{\bar{x}_{q'}}\, x_{q'} \quad \text{...(cvii)}.$$

Having reached this result numerically, it can easily be proved algebraically*. For

$$r_{qq'} = \frac{\Gamma(q')}{\Gamma(q)}\,\frac{\Gamma(q+\lambda)}{\Gamma(q'+\lambda)}\,\frac{\sigma_{q'}}{\sigma_q},$$

and the terms in q', $q = \bar{x}_q/\bar{x}_{q'}$; hence

$$\begin{aligned} \tilde{x}_q - \bar{x}_q &= r_{qq'}\frac{\sigma_q}{\sigma_{q'}}(x_{q'} - \bar{x}_{q'}) \\ &= \frac{\bar{x}_q}{\bar{x}_{q'}}(x_{q'} - \bar{x}_{q'}), \end{aligned}$$

or,

$$\tilde{x}_q = \frac{\bar{x}_q}{\bar{x}_{q'}}\, x_{q'}.$$

This form of relationship is free from the parameters of the parent curve, and suggests that when we increase the total range of the sample we shall on the average increase all the subranges in the same proportion. How far this property extends beyond the curve types on the (β_1, β_2) biquadratic I am not at present able to state. The difficulty arises from the need of a fixed terminal to measure $\bar{x}_q$ from. It would be interesting, however, to investigate further the distribution of rank-intervals in the case of other curves.

* I might have removed the numerical road to the theorem, but I think it may suggest to others that mere computing work may occasionally lead one to analytical theorems.

Still dealing with the same numerical example, i.e. $y = y_0 (a/x)^{\cdot 8}$, we may indicate how the correlation is determined between any two rank-intervals; let, say, the first and last,

$$r_{x_2-x_1,\, x_{11}-x_{10}} = \frac{\{(x_2 - x_1)(x_{11} - x_{10})\} - (\bar{x}_2 - \bar{x}_1)(\bar{x}_{11} - \bar{x}_{10})}{\sigma_{x_2-x_1} \times \sigma_{x_{11}-x_{10}}}.$$

Now $$\{(x_2 - x_1)(x_{11} - x_{10})\} = \{x_2 x_{11}\} + \{x_1 x_{10}\} - \{x_1 x_{11}\} - \{x_2 x_{10}\}.$$

Substituting the values from pp. 244 and 245 we find

$$r_{x_2-x_1,\, x_{11}-x_{10}} = -\cdot 050{,}268.$$

The negative correlation is due I think to the fact that the total range being limited, an increased first interrank must denote a decreased last interrank interval. The two intervals, however, being far apart, there is only a very small correlation between their magnitudes. The correctness of this explanation will be confirmed as we take further curves round the biquadratic.

We may illustrate further our theory of the ranking of the biquadratic curves by considering on the same curve the problem of total range distribution. The only expression I have been able to find for the frequency curve of range R is of the form*

$$y = y_0 \left(\frac{R}{a}\right)^{n(1-m)-1} \mathop{S}_{t=0}^{\infty} c_t \left(1 - \frac{R}{a}\right)^t \text{.....................(cviii)},$$

* It is, perhaps, better retained in the definite integral form, which exhibits its relation to the known form in the case of a rectangular parent population.

The chance of x_1 and x_n occurring with the values a_{x_1} and a_{x_n} is

$$P = \frac{(n)!}{(n-2)!}\, da_{x_1} (a_{x_n} - a_{x_1})^{n-2}\, da_{x_n}$$
$$= \frac{n(n-1)(1-m)^2}{a^2} \left(\frac{x_1}{a}\right)^{-m} dx_1 \left(\left(\frac{x_n}{a}\right)^{1-m} - \left(\frac{x_1}{a}\right)^{1-m}\right)^{n-2} \left(\frac{x_n}{a}\right)^{-m} dx_n.$$

Put $R = x_n - x_1$, eliminate x_n and integrate for all possible values of x_1, i.e. from 0 to $a - R$, then we find the chance of a range lying between R and $R + dR$ is

$$P = \frac{n(n-1)(1-m)^2}{a^2} \int_0^{a-R} \left(\frac{x}{a}\right)^{-m} \left(\left(\frac{x+R}{a}\right)^{1-m} - \left(\frac{x}{a}\right)^{1-m}\right)^{n-2} \left(\frac{x+R}{a}\right)^{-m} dx\, dR.$$

Write $z = \left(\dfrac{x}{x+R}\right)^{1-m}$; we obtain after some reductions

$$P = \frac{n(n-1)(1-m)}{a} \left(\frac{R}{a}\right)^{n(1-m)-1} \int_0^{\left(1-\frac{R}{a}\right)^{1-m}} \frac{(1-z)^{n-2}\, dz}{(1 - z^{\frac{1}{1-m}})^{n(1-m)}}\, dR.$$

Write $z = \left(1 - \dfrac{R}{a}\right)^{1-m} u$, then the frequency curve for the distribution of N ranges is

$$y = \frac{Nn(n-1)(1-m)}{a} \left(\frac{R}{a}\right)^{n(1-m)-1} \left(1 - \frac{R}{a}\right)^{1-m} \int_0^1 \frac{\left(1 - \left(1 - \frac{R}{a}\right)^{1-m} u\right)^{n-2}}{\left(1 - \left(1 - \frac{R}{a}\right) u^{\frac{1}{1-m}}\right)^{n(1-m)}}\, du \text{ ......(cix).}$$

This is expansible in a series like (cviii) above.

Put $m = 0$ and the integral becomes

$$\int_0^1 \left(1 - \left(1 - \frac{R}{a}\right) u\right)^{-2} du = \frac{1}{1 - \frac{R}{a}} \left[\frac{1}{1 - \left(1 - \frac{R}{a}\right) u}\right]_0^1 = \frac{1}{1 - \frac{R}{a}} \left(\frac{a}{R} - 1\right),$$

or $$y = \frac{Nn(n-1)}{a} \left(\frac{R}{a}\right)^{n-2} \left(1 - \frac{R}{a}\right) \text{ ....(cx),}$$

where c_t is a function of n, m and t. For the series I have not been able to find any generative function. The graph of the curve of range-frequency could only be constructed with much labour from this formula. Consequently I have had to fall back on the determination of the β's of the range distribution, and from them take the corresponding frequency curve. This can afterwards be compared for a certain number of selected points with the results obtained from (cix) by quadratures.

For the curve $y = \dfrac{N}{5a}\left(\dfrac{a}{x}\right)^{\cdot 8}$, we have by the formula (lxxiii), assuming $a = 1$,

$$\bar{x}_1 = \{x_1\} = \cdot 0002{,}2893{,}7729, \quad \bar{x}_{11} = \{x_2\} = \cdot 6875,$$

$$\mu_1' = \bar{x}_{11} - \bar{x}_1 = \cdot 6872{,}7106{,}2271,$$

$$\{x_1^2\} = \cdot 0000{,}0283{,}5142, \quad \{x_1 x_{11}\} = \cdot 0001{,}7442{,}8746, \quad \{x_{11}^2\} = \cdot 5238{,}0952{,}8746,$$

giving
$$\mu_2' = \{(x_{11} - x_1)^2\} = \cdot 5234{,}6350{,}1460.$$

Again,
$$\{x_1^3\} = \cdot 0000{,}0012{,}9430, \quad \{x_{11}^3\} = \cdot 4230{,}7692{,}3077,$$

$$\{x_1^2 x_{11}\} = \cdot 0000{,}0228{,}9923, \quad \{x_1 x_{11}^2\} = \cdot 0001{,}4088{,}4756,$$

giving
$$\mu_3' = \{(x_{11} - x_1)^3\} = \cdot 4226{,}6100{,}9148.$$

corresponding, after change of a for b, and after change of origin, with the value given in *Biometrika*, Vol. XXIII. p. 398. This is the range curve in the case of a rectangular population.

Another fairly easy solution can be obtained for this range frequency curve in the special case of the curve $y = y_0\sqrt{\dfrac{a}{x}}$. Here $m = \cdot 5$, and consequently the range has for its frequency distribution

$$y = \frac{Nn(n-1)}{2a}\left(\frac{R}{a}\right)^{\frac{1}{2}n-1}\left(1-\frac{R}{a}\right)^{\frac{1}{2}}\int_0^1 \frac{\left(1 - u\sqrt{1-\dfrac{R}{a}}\right)^{n-2}}{\left(1 - u^2\left(1-\dfrac{R}{a}\right)\right)^{\frac{1}{2}n}}\,du.$$

Take
$$\sqrt{1-\frac{R}{a}}\,u = \frac{1-w}{1+w} \quad\text{or}\quad w = \frac{1-\sqrt{1-\dfrac{R}{a}}\,u}{1+\sqrt{1-\dfrac{R}{a}}\,u}.$$

Then
$$\sqrt{1-\frac{R}{a}}\,du = -\frac{2}{(1+w)^2}\,dw,$$

and for the limits of the integral we have

$$u = 0,\ w = 1,\ \text{and}\ u = 1,\ w = \frac{1-\sqrt{1-\dfrac{R}{a}}}{1+\sqrt{1-\dfrac{R}{a}}} = w_0,\ \text{say}.$$

Hence
$$y = \frac{Nn(n-1)}{4a}\left(\frac{R}{a}\right)^{\frac{1}{2}n-1}\int_{w_0}^1 w^{\frac{1}{2}n-2}\,dw$$

$$\frac{Nn(n-1)}{2a(n-2)}\left(\frac{R}{a}\right)^{\frac{1}{2}n-1}\left(1 - \left(\frac{1-\sqrt{1-\dfrac{R}{a}}}{1+\sqrt{1+\dfrac{R}{a}}}\right)^{\frac{1}{2}n-1}\right)$$

or finally,
$$y = \frac{Nn(n-1)}{2a(n-2)}\left(\left(\sqrt{\frac{R}{a}}\right)^{n-2} - \left(1-\sqrt{1-\frac{R}{a}}\right)^{n-2}\right) \quad \text{......................(cxi)},$$

a curve of range a, and easily graphed.

For the fourth moment

$\{x_1^4\} = \cdot 0000{,}0001{,}1810$, $\{x_1^2 x_{11}^2\} = \cdot 0000{,}0192{,}0580$, $\{x_{11}^4\} = \cdot 3548{,}3870{,}9677$,

$\{x_1^3 x_{11}\} = \cdot 0000{,}0010{,}8554^5$, $\{x_1 x_{11}^3\} = \cdot 0001{,}1816{,}1408^5$,

and accordingly $\mu_4' = \{(x_{11} - x_1)^4\} = \cdot 3543{,}7716{,}5114.$

Lastly for the fifth moment

$\{x_1^5\} = \cdot 0000{,}0000{,}0208$, $\{x_1^4 x_{11}\} = \cdot 0000{,}0001{,}0170$, $\{x_1^3 x_{11}^2\} = \cdot 0000{,}0009{,}3477^5$,

$\{x_1^2 x_{11}^3\} = \cdot 0000{,}0165{,}3833$, $\{x_1 x_{11}^4\} = \cdot 0001{,}0175{,}0102$, $\{x_{11}^5\} = \cdot 3055{,}5555{,}5556$,

whence $\mu_5' = \{(x_{11} - x_1)^5\} = \cdot 3050{,}6287{,}5781.$

Transferring to the mean, we find in the usual way

$$\mu_2 = \cdot 0511{,}2198{,}8425, \quad \mu_3 = -\cdot 0073{,}6963{,}4,$$
$$\mu_4 = \cdot 0066{,}4817{,}9, \quad \mu_5 = -\cdot 0022{,}6303{,}7.$$

And thus we obtain, on reinstating a*,

$$\sigma_R = \cdot 2261{,}0172a, \quad \beta_1 = \cdot 406{,}5076, \quad \beta_2 = 2\cdot 543{,}825^-.$$

If we consider what form a Pearson curve would take with these values of β_1, β_2, we find it to be of Type I, a limited range curve, and its equation is

$$y = y_0 \left(1 + \frac{x}{\cdot 999{,}389a}\right)^{1\cdot 201{,}293} \left(1 - \frac{x}{\cdot 000{,}526}\right)^{\cdot 000{,}632} \quad \text{............(cxii).}$$

The range is thus $\cdot 999{,}915a$, only differing in the fifth place of decimals from a, and its mean is at $\cdot 687{,}432a$, while the true mean is at $\cdot 687{,}271a$. It would seem accordingly that we should obtain as good a fit by determining our curve from mean, range, and standard deviation, as by using the first four moments. The method of fitting is then very easy†. We have

$$y = 2\cdot 176{,}046\,\frac{N}{a} \left(1 + \frac{x}{\cdot 998{,}2886a}\right)^{1\cdot 202{,}187} \left(1 - \frac{x}{\cdot 001{,}7114}\right)^{\cdot 002{,}061} \quad \text{...(cxiii).}$$

The constants clearly do not differ much from those of (cxii), but a better comparison can be obtained by comparing the moment coefficients about the end of the range a.

	Actual Values	Curve (cxii)	Curve (cxiii)
Mean (μ_1')	$\cdot 687{,}271a$	$\cdot 687{,}432a$	$\cdot 687{,}271a$
Range	a	$\cdot 999{,}915a$	a
μ_2'	$\cdot 5234{,}6350a^2$	$\cdot 5234{,}6350a^2$	$\cdot 5234{,}6350a^2$
μ_3'	$\cdot 4226{,}6101a^3$	$\cdot 4226{,}6101a^3$	$\cdot 4226{,}7231a^3$
μ_4'	$\cdot 3543{,}7717a^4$	$\cdot 3543{,}7717a^4$	$\cdot 3544{,}0563a^4$
μ_5'	$\cdot 3050{,}6288a^5$	$\cdot 3051{,}4018a^5$	$\cdot 3051{,}1026a^5$

The differences are practically very slight, but the curve (cxiii) is nearer in three of the seven quantities tabled to the true values than the curve (cxii), and (cxii) is only

* The reader must remember that the relation between the σ of the parent population and a is given by (lvii), or $a = \sigma 1\cdot 2\sqrt{11} = \sigma \times 3\cdot 9799{,}4976$, or $\sigma = \cdot 2512{,}5945a$, and $\sigma_R = \cdot 8998{,}7349\sigma$, or about nine-tenths of the parent population standard deviation.

† See *Phil. Trans.* Vol. 186A, pp. 370—371.

nearer in two of the constants. The relative ease with which it can be computed suggests that it is the better method of fitting the distribution of range. How does it compare, however, with the true curve (cix)?

I owe to my colleague, Mr E. C. Fieler, the system of points obtained by quadrature for the curve in (cix), and to my colleague, Miss Brenda Stoessiger, the ordinates for plotting (cxii). See Diagram III, p. 252. The elasticity of Type I is remarkable; it is here almost a triangle; but the points of the integral curve (cix) to the third decimal show agreement with (cxii), and it is not feasible to indicate any divergence even on the original diagram, which measured $10'' \times 4{\cdot}5''$. This is very good evidence that the range curve may be adequately obtained from range, mean, and standard deviation, without using the troublesome quadratures necessary in the case of (cix). The accordance is especially noteworthy considering the smallness of the sample.

We may now proceed to test how far it is feasible to obtain the frequency-distribution of any rank-interval, $i_{q'-q}$, between the q'th and qth ranks. We will simply use the letter i for brevity during the analysis, i.e. $i = x_{q'} - x_q$. The chance of a value of α_x falling in $d\alpha_x$, and a value of $\alpha_{x'}$ in $d\alpha_{x'}$

$$= \frac{n!}{(q-1)!\,(q'-q-1)!\,(n-q')!}\,\alpha_x^{q-1}\,d\alpha_x\,(\alpha_{x'}-\alpha_x)^{q'-q-1}\,d\alpha_{x'}\,(1-\alpha_{x'})^{n-q'}.$$

Hence in the case of the curve $y = y_0\,(x/a)^{-m}$, which gives $\alpha_x = (x/a)^{1-m}$ and $d\alpha_x = \left(\frac{1-m}{a}\right)(x/a)^{-m}$, the chance of x_q lying in dx_q and $x_{q'}$ in $dx_{q'}$

$$= \frac{n!\,(1-m)^2}{a^2(q-1)!\,(q'-q-1)!\,(n-q')!}\left(\frac{x_q}{a}\right)^{q(1-m)-1}\left(\left(\frac{x_{q'}}{a}\right)^{1-m}-\left(\frac{x_q}{a}\right)^{1-m}\right)^{q'-q-1}$$
$$\left(1-\left(\frac{x_{q'}}{a}\right)^{1-m}\right)^{n-q'}\left(\frac{x_{q'}}{a}\right)^{-m} dx_q\,dx_{q'}.$$

Let us write $x_{q'} = x_q + i$ and change our variables to x_q and i, then the value of i remaining constant, we need to integrate x_q from 0 to $a - i$ in order to obtain the distribution curve of i. We will write $x_q/a = x'$, $i/a = i'$ and

$$\chi = \frac{n!\,(1-m)^2}{a\,(q-1)!\,(q'-q-1)!\,(n-q')!}.$$

Then dropping the di' we have, for the distribution curve of i',

$$y = \chi\int_0^{1-i'}(x')^{q(1-m)-1}\left((x'+i')^{1-m}-(x')^{1-m}\right)^{q'-q-1}\left(1-(x'+i')^{1-m}\right)^{n-q'}(x'+i')^{-m}\,dx'.$$

Write $x' = i'z/(1-z)$, and we find

$$y = \chi i'^{\,q'(1-m)-1}\int_0^{1-i'} z^{q(1-m)-1}\,\frac{(1-z^{1-m})^{q'-q-1}}{(1-z)^{n(1-m)}}\left((1-z)^{1-m}-i'^{\,(1-m)}\right)^{n-q'}dz \qquad \text{.........(cxiv).}$$

This is the curve of distribution for the interrank interval $i_{q'-q} = ai'$. It does not seem feasible to reduce it substantially in form except for special values of q and q'; for example, in the case of the total range when $q = 1$, $q' = n$; see p. 248 above.

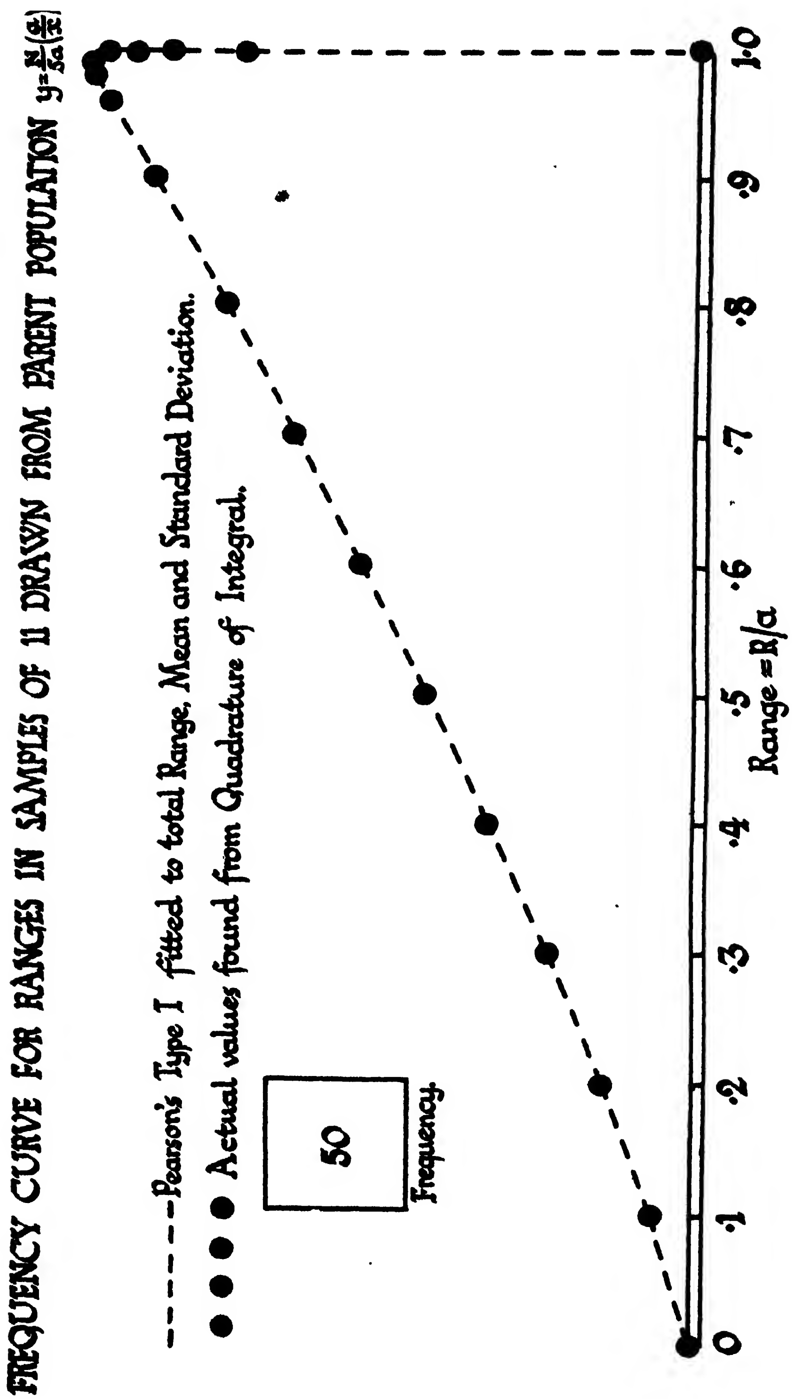

FREQUENCY CURVE FOR RANGES IN SAMPLES OF 11 DRAWN FROM PARENT POPULATION y=N/5a(x/a)
- - - - Pearson's Type I fitted to total Range, Mean and Standard Deviation.
● ● ● ● Actual values found from Quadrature of Integral.
50
Frequency
0
·1
·2
·3
·4
·5
·6
·7
·8
·9
1·0
Range = R/a

DIAGRAM III.

Let $$w_{n-q'} = \int_0^{1-i'} z^{q(1-m)-1} \frac{(1-z^{1-m})^{q'-q-1}}{(1-z)^{n(1-m)}} ((1-z)^{1-m} - (i')^{1-m})^{n-q'} dz,$$

then $w_{n-q'} = 0$ when $i = 1$, and $\frac{dw_p}{di'} = 0$ when $z = 1 - i'$, as long as $p > 1$.

We can now consider the moment coefficients of (cxiv) about the end of the range a of the parent population, remembering that $i = ai'$.

We have $$\mu_s' = \int_0^1 yi^s di = \chi a^{s+1} \int_0^1 i'^{q'(1-m)-1+s} w_{n-q} di',$$
or, putting $i'^{(1-m)} = v$,

$$\mu_s' = \frac{\chi}{1-m} a^{s+1} \int_0^1 v^{q'-1+\frac{s}{1-m}} w'_{n-q'} dv,$$

where $$w'_{n-q'} = \int_0^{1-v^{\frac{1}{1-m}}} \frac{z^{q(1-m)-1}(1-z^{1-m})^{q'-q-1}}{(1-z)^{m(1-m)}} ((1-z)^{1-m} - v)^{n-q'} dz.$$

Integrate by parts and we find the part outside the integral vanishes when $v = 0$ and $v = 1$.

Hence $$\mu_s' = \frac{\chi}{1-m} a^{s+1} (-1) \int_0^1 \frac{v^{q'+\frac{s}{1-m}}}{q' + \frac{s}{1-m}} \frac{dw'_{n-q'}}{dv} dv.$$

Now $\frac{dw'_{n-q'}}{dv}$ consists of two parts, that arising from the differentiation of the upper limit, and that from the differentiation of v in the integrand. The former vanishes, and the latter is $-(n-q')\, w_{n-q'-1}$. Thus we have

$$\mu_s' = \frac{\chi}{1-m} a^{s+1} \frac{n-q'}{q' + \frac{s}{1-m}} \int_0^1 v^{q'+\frac{s}{1-m}} w_{n-q'-1} di'.$$

We can continuously repeat this process until the power of $((1-z)^{\frac{1}{1-m}} - v)$ in the z-integral is exhausted, and we then find

$$\mu_s' = \frac{\chi}{1-m} a^{s+1} \frac{(n-q')!}{\left(q' + \frac{s}{1-m}\right)\left(q' + 1 + \frac{s}{1-m}\right) \dots \left(n - 1 + \frac{s}{1-m}\right)} \int_0^1 v^{n-1+\frac{s}{1-m}} w_0 dv.$$

We now once more integrate by parts, but the integrand in w_0 no longer contain v, only the upper limit is a function of v. The part between limits $\left[v^{n+\frac{s}{1-m}} w_0\right]_0^1$ still vanishes at both limits, and thus we reach

$$\mu_s' - \frac{\chi}{1-m} a^{s+1} \frac{\Gamma\left(q' + \frac{s}{1-m}\right) \Gamma(n - q' + 1)}{\Gamma\left(n + 1 + \frac{s}{1-m}\right)} \int_0^1 v^{n+\frac{s}{1-m}} \left(-\frac{dw_0}{dv}\right) dv,$$

where $$w_0 = \int_0^{1-v^{\frac{1}{1-m}}} \frac{z^{q(1-m)-1}(1-z^{1-m})^{q'-q-1}}{(1-z)^{n(1-m)}} dz,$$

and accordingly

$$\frac{dw_0}{dv} = -\frac{1}{1-m}\frac{v^{\frac{m}{1-m}}(1-v^{\frac{1}{1-m}})^{q(1-m)-1}}{v^n}(1-(v^{\frac{1}{m-1}})^{1-m})^{q'-q-1}.$$

Thus we have

$$\mu_s' = \frac{\chi a^{s+1}}{(1-m)^2}\frac{\Gamma\left(q'+\frac{s}{1-m}\right)\Gamma(n-q'-1)}{\Gamma\left(n+1+\frac{s}{1-m}\right)}\int_0^1 v^{\frac{s+m}{1-m}}(1-v^{\frac{1}{1-m}})^{q(1-n}$$

$$\times(1-(1-v^{\frac{1}{1-m}})^{1-m})^{q'-q-1}\,dv.$$

Take $$u=(1-v^{\frac{1}{1-m}})^{1-m},$$

$$du = -v^{\frac{m}{1-m}}(1-v^{\frac{1}{1-m}})^{-m}\,dv,$$

and when $v=0$, $u=1$, and $v=1$, $u=0$. Thus we have

$$\mu_s' = \frac{\chi a^{s+1}}{(1-m)^2}\frac{\Gamma\left(q'+\frac{s}{1-m}\right)\Gamma(n-q'+1)}{\Gamma\left(n+1+\frac{s}{1-m}\right)}\int_0^1(1-u^{\frac{1}{1-m}})^s u^{q-1}(1-u)^{q'-q-1}\,du.$$

Substituting the value of χ, we have

$$\mu_s' = a^s\frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{s}{1-m}\right)}\frac{\Gamma\left(q'+\frac{s}{1-m}\right)}{\Gamma(q')}\frac{\Gamma(q')}{\Gamma(q)\Gamma(q'-q)}$$

$$\times\left(\frac{\Gamma(q)\Gamma(q'-q)}{\Gamma(q')}-\frac{s}{1!}\frac{\Gamma\left(q+\frac{1}{1-m}\right)\Gamma(q'-q)}{\Gamma\left(q'+\frac{1}{1-m}\right)}+\frac{s(s-1)}{2!}\frac{\Gamma\left(q+\frac{2}{1-m}\right)\Gamma(q'-q)}{\Gamma\left(q'+\frac{2}{1-m}\right)}\right.$$

$$\left.-\frac{s(s-1)(s-2)}{3!}\frac{\Gamma\left(q+\frac{3}{1-m}\right)\Gamma(q-q)}{\Gamma\left(q'+\frac{3}{1-m}\right)}+\dots\right)\quad\text{.........(cxv).}$$

Or, symbolically,

$$\mu_s' = a^s\frac{\Gamma(n+1)}{\Gamma(q')}\frac{\Gamma\left(q'+\frac{s}{1-m}\right)}{\Gamma\left(n+1+\frac{s}{1-m}\right)}(1-E)^s\frac{\Gamma\left(q+\frac{0}{1-m}\right)}{\Gamma\left(q'+\frac{0}{1-m}\right)}\quad\text{...............(cxv}^{\text{bis}}\text{).}$$

If $s=0$, we have

$\mu_0'=1$.

If $s=1$:

$$\mu_1' = \frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{1}{m+1}\right)}\left(\frac{\Gamma\left(q'+\frac{1}{m}\right)}{\Gamma(q')}-\frac{\Gamma\left(q+\frac{1}{m}\right)}{\Gamma(q)}\right)\quad\text{.........................(cxvi).}$$

If $s = 2$:

$$\mu_2' = \frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{2}{1-m}\right)}\left\{\frac{\Gamma\left(q'+\frac{2}{1-m}\right)}{\Gamma(q')} - 2\frac{\Gamma\left(q'+\frac{2}{1-m}\right)}{\Gamma\left(q'+\frac{1}{1-m}\right)}\frac{\Gamma\left(q+\frac{1}{1-m}\right)}{\Gamma(q)} + \frac{\Gamma\left(q+\frac{2}{1-m}\right)}{\Gamma(q)}\right\} \quad \text{...(cxvii).}$$

If $s = 3$:

$$\mu_3' = \frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{3}{1-m}\right)}\left\{\frac{\Gamma\left(q'+\frac{3}{1-m}\right)}{\Gamma(q')} - 3\frac{\Gamma\left(q'+\frac{3}{1-m}\right)}{\Gamma\left(q'+\frac{1}{1-m}\right)}\frac{\Gamma\left(q+\frac{1}{1-m}\right)}{\Gamma(q)} + 3\frac{\Gamma\left(q'+\frac{3}{1-m}\right)}{\Gamma\left(q'+\frac{2}{1-m}\right)}\frac{\Gamma\left(q+\frac{2}{1-m}\right)}{\Gamma(q)} - \frac{\Gamma\left(q+\frac{3}{1-m}\right)}{\Gamma(q)}\right\} \quad \text{...(cxviii).}$$

If $s = 4$:

$$\mu_4' = \frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{4}{1-m}\right)}\left(\frac{\Gamma\left(q'+\frac{4}{1-m}\right)}{\Gamma(q')} - 4\frac{\Gamma\left(q'+\frac{4}{1-m}\right)}{\Gamma\left(q'+\frac{1}{1-m}\right)}\frac{\Gamma\left(q+\frac{1}{1-m}\right)}{\Gamma(q)} + 6\frac{\Gamma\left(q'+\frac{4}{1-m}\right)}{\left(q'+\frac{2}{1-m}\right)}\frac{\Gamma\left(q+\frac{2}{1-m}\right)}{\Gamma(q)} - 4\frac{\Gamma\left(q'+\frac{4}{1-m}\right)}{\Gamma\left(q'+\frac{3}{1-m}\right)}\frac{\Gamma\left(q+\frac{3}{1-m}\right)}{\Gamma(q)} + \frac{\Gamma\left(q+\frac{4}{1-m}\right)}{\Gamma(q)}\right\} \quad \text{...(cxix),}$$

and so on.

The complete Γ-functions can be found by aid of the tables and thus the moment coefficients about $i_{q'-q} = 0$ calculated. These can be transferred to the mean, and $\sigma_{q'-q}$, β_1 and β_2 ascertained and then a Pearson curve fitted. There is little doubt that it will give as satisfactory results as curves thus deduced for the entire range.

It is clear that the forms we get for μ's are precisely what we obtain by expanding $(x_{q'} - x_q)^s$ and using (lxxix). The only interest of the investigation is the discovery of the intractable distribution curve (cxiv), and the transition to the manageable moments of that curve.

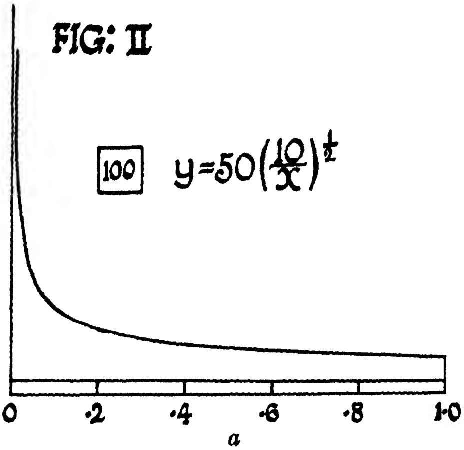
Illustration II. Type VIII.

$$\text{Curve (ii)}: \; y = \frac{N}{2a}\sqrt{\frac{a}{x}},$$

$$\sigma = \cdot 2981{,}4240a, \quad \beta_1 = 1 \cdot 428{,}571, \quad \beta_2 = 2 \cdot 142{,}857.$$

This curve is given in Fig. II.

FIG: II



Here by (lxxvi) and (lxxvii) we have ratio of first to second prizes 21 to 10, and ratio of first and second penalties 5 to 3. It will thus appear that the Galton Ratios change considerably with the nature of the parent population, e.g.

Parent Population	Prize Ratio	Penalty Ratio
$y=\frac{N}{5a}\left(\frac{a}{x}\right)^{\cdot 8}$	24 to 10	13 to 10 (4 to 3)
$y=\frac{N}{2a}\left(\frac{a}{x}\right)^{\cdot 5}$	21 to 10	17 to 10 (5 to 3)
Rectangle $y=\frac{N}{a}$	20 to 10	20 to 10 (6 to 3)
Exponential Curve	30 to 10	19 to 10
Normal Curve	*circa* 26 to 10	26 to 10

Further, the penalty ratio* is less than the prize ratio in all cases, where there is lack of symmetry.

We have the following further results:

Mean Rank Variates	Standard Deviation of Rank Variates
$\bar{x}_1 = \cdot 0128{,}2051{,}28a$,	$\sigma_{x_1} = \cdot 0238{,}3768{,}41a$,
$\bar{x}_2 = \cdot 0384{,}6153{,}85a$,	$\sigma_{x_2} = \cdot 0467{,}3022{,}28a$,
$\bar{x}_3 = \cdot 0769{,}2307{,}69a$,	$\sigma_{x_3} = \cdot 0712{,}1693{,}08a$,
$\bar{x}_4 = \cdot 1282{,}0512{,}82a$,	$\sigma_{x_4} = \cdot 0959{,}3993{,}30a$,
$\bar{x}_5 = \cdot 1923{,}0769{,}23a$,	$\sigma_{x_5} = \cdot 1195{,}8178{,}27a$,
$\bar{x}_6 = \cdot 2692{,}3076{,}92a$,	$\sigma_{x_6} = \cdot 1407{,}9234{,}79a$,
$\bar{x}_7 = \cdot 3589{,}7435{,}90a$,	$\sigma_{x_7} = \cdot 1580{,}6189{,}75a$,
$\bar{x}_8 = \cdot 4615{,}3846{,}15a$,	$\sigma_{x_8} = \cdot 1695{,}3020{,}47a$,
$\bar{x}_9 = \cdot 5769{,}2307{,}69a$,	$\sigma_{x_9} = \cdot 1726{,}1844{,}04a$,
$\bar{x}_{10} = \cdot 7051{,}2820{,}51a$,	$\sigma_{x_{10}} = \cdot 1631{,}0652{,}40a$,
$\bar{x}_{11} = \cdot 8461{,}5384{,}61a$,	$\sigma_{x_{11}} = \cdot 1317{,}4597{,}49a$.

* The penalty ratio is taken from the mediocrity end of the skew curves, or from those less than mediocrity.

We see in the case of this curve the absolute variability reaching a maximum for the variability of the 9th of the eleven ranks, whereas in the previous case it reached its maximum in the tenth rank only. On the other hand, the relative variation decreases from the lowest to the highest rank and, although the size of the sample is the same, is much less than in Curve (i). Thus:

$$\begin{array}{ll} V_1 = 185{\cdot}934, & V_8 = 36{\cdot}521, \\ V_2 = 121{\cdot}149, & V_9 = 29{\cdot}921, \\ V_3 = 92{\cdot}582, & V_{10} = 23{\cdot}131, \\ V_4 = 74{\cdot}833, & V_{11} = 15{\cdot}570. \end{array}$$

Passing to the Rank-Intervals we have

$$\begin{array}{ll}
\bar{x}_2 - \bar{x}_1 = {\cdot}0256{,}4102{,}57a, & \sigma_{x_2 - x_1} = {\cdot}0360{,}0194{,}10a, \\
\bar{x}_3 - \bar{x}_2 = {\cdot}0384{,}6153{,}84a, & \sigma_{x_3 - x_2} = {\cdot}0467{,}3022{,}28a, \\
\bar{x}_4 - \bar{x}_3 = {\cdot}0512{,}8205{,}13a, & \sigma_{x_4 - x_3} = {\cdot}0568{,}4150{,}84a, \\
\bar{x}_5 - \bar{x}_4 = {\cdot}0641{,}0256{,}41a, & \sigma_{x_5 - x_4} = {\cdot}0666{,}1733{,}88a, \\
\bar{x}_6 - \bar{x}_5 = {\cdot}0769{,}2307{,}69a, & \sigma_{x_6 - x_5} = {\cdot}0769{,}2307{,}69a, \\
\bar{x}_7 - \bar{x}_6 = {\cdot}0897{,}4358{,}97a, & \sigma_{x_7 - x_6} = {\cdot}0856{,}1952{,}75a, \\
\bar{x}_8 - \bar{x}_7 = {\cdot}1025{,}6410{,}25a, & \sigma_{x_8 - x_7} = {\cdot}0949{,}5590{,}75a, \\
\bar{x}_9 - \bar{x}_8 = {\cdot}1153{,}8461{,}54a, & \sigma_{x_9 - x_8} = {\cdot}1042{,}2194{,}95a, \\
\bar{x}_{10} - \bar{x}_9 = {\cdot}1282{,}0512{,}82a, & \sigma_{x_{10} - x_9} = {\cdot}1134{,}3489{,}07a, \\
\bar{x}_{11} - \bar{x}_{10} = {\cdot}1410{,}2564{,}11a, & \sigma_{x_{11} - x_{10}} = {\cdot}1226{,}0664{,}49a.
\end{array}$$

Now a very remarkable result flows from the intervals, the difference of the intervals—the second difference of the rank variates is constant $= {\cdot}0128{,}2051{,}28$. Is this peculiar to this curve, or a more general result? A very brief investigation shows that it is not a general principle, but the fact for this case suggested an inquiry into the differences of the mean variates of ranks. Returning to the formula (lxxiii) we have

$$\bar{x}_{q+1} - \bar{x}_q = \frac{a\Gamma(n+1)}{\Gamma\left(n+1+\frac{1}{1-m}\right)} \left(\frac{\Gamma\left(q+1+\frac{1}{1-m}\right)}{\Gamma(q+1)} - \frac{\Gamma\left(q+\frac{1}{1-m}\right)}{\Gamma(q)} \right),$$

or

$$\Delta\bar{x}_q = \frac{a\Gamma(n+1)}{\left(n+1+\frac{1}{1-m}\right)} \frac{\Gamma\left(q+\frac{1}{1-m}\right)}{\Gamma(q+1)} \left(\frac{1}{1-m} - 0\right),$$

$$\Delta^2\bar{x}_q = \frac{a\Gamma(n+1)}{\Gamma\left(n+1+\frac{1}{1-m}\right)} \frac{\Gamma\left(q+\frac{1}{1-m}\right)}{\Gamma(q+2)} \left(\frac{1}{1-m} - 0\right)\left(\frac{1}{1-m} - 1\right),$$

and continuing

$$\Delta^s\bar{x}_q = a\frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{1}{1-m}\right)} \frac{\Gamma\left(q+\frac{1}{1-m}\right)}{\Gamma(q+s)} \left(\frac{1}{1-m}\right)\left(\frac{1}{1-m} - 1\right)\dots\left(\frac{1}{1-m} - (s-1)\right)$$

.........(cxx).

Therefore when $s = 1 + \dfrac{1}{1-m}$, the sth difference of the mean rank-variates—the $s-1$ difference of the mean rank-intervals—will vanish. When $m = \cdot 5$, $s = 3$, or the third mean rank-variate difference, the second mean rank-interval difference is zero, as we have just noted. When $m = 0$, $\Delta^2 x_q = 0$, or for a rectangular parent population all the mean rank-intervals are equal. Again, if $m = \cdot 8$, the sixth mean rank-variate differences are zero, or the fourth mean interrank interval differences are constant. Similar properties hold for the differences of the $\mu_t'(n, q)$ moment coefficients, but it is doubtful whether they would be of service for computing purposes.

Returning to the results of the rank-intervals on p. 257 we see a marked change in the ratio of the first to the last; it now lies between 6 and 7 instead of amounting to 200, as in the case of curve (i); thus there is an approach to the equality of rank-intervals, which comes with the rectangle. A similar remark applies to the absolute variabilities of the last and first intervals. For curve (i) their ratio was 39·5, while for the present curve it is only 3·4.

Our results for σ_{x_q} and $\sigma_{x_{q'}-x_q}$ enable us to obtain the following system of correlations:

$$
\begin{array}{ll}
r_{x_1,x_2} = \cdot 6534{,}5026, & r_{x_1,x_3} = \cdot 4979{,}2960, \\
r_{x_2,x_3} = \cdot 7620{,}0077, & r_{x_1,x_4} = \cdot 4024{,}7170, \\
r_{x_3,x_4} = \cdot 8082{,}9038, & \ldots\ldots\ldots\ldots\ldots\ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots & r_{x_1,x_7} = \cdot 2368{,}1276, \\
r_{x_7,x_8} = \cdot 8342{,}1006, & r_{x_1,x_8} = \cdot 1975{,}5159, \\
r_{x_8,x_9} = \cdot 8145{,}7314, & r_{x_1,x_9} = \cdot 1609{,}2022, \\
r_{x_9,x_{10}} = \cdot 7730{,}9696, & r_{x_1,x_{10}} = \cdot 1244{,}0692, \\
r_{x_{10},x_{11}} = \cdot 6731{,}0804, & r_{x_1,x_{11}} = \cdot 0837{,}3929.
\end{array}
$$

These values, all found directly, confirm the general theory that if $q < q' < q''$, then $r_{qq''} = r_{qq'} r_{q'q''}$. We see further that the correlation between the variates of adjacent ranks reaches its maximum not at the terminals but towards the centre of the system. The second column indicates the reduction in association as our ranks are farther and farther apart—the variates of the first and second ranks have almost eight times as close an association as those of the first and last ranks. Lastly, we may note that the tendency for the correlations in the first column to become a symmetrical system indicates that our curve is approaching the rectangular parent population.

We have the following system of correlations for interrank distances and interrank distances with rank-variates:

$$
\begin{array}{ll}
r_{x_2-x_1,\,x_3-x_2} = \cdot 1395{,}6901, & r_{x_2,\,x_2-x_1} = \cdot 6593{,}8048, \\
r_{x_2-x_1,\,x_4-x_3} = \cdot 0734{,}3468, & r_{x_3,\,x_4-x_3} = \cdot 1113{,}6920, \\
r_{x_3-x_2,\,x_4-x_3} = \cdot 0848{,}6342, & r_{x_3,\,x_3-x_2} = \cdot 7620{,}0076,
\end{array}
$$

$r_{x_2-x_1,\,x_6-x_5} = -\cdot 0068{,}5631$ (shows change of sign occurs about median interval),

$$
\begin{array}{lll}
r_{x_2-x_1,\,x_8-x_7} = -\cdot 0549{,}4837, & r_{x_3-x_2,\,x_8-x_7} = -\cdot 0635{,}0006, & r_{x_4-x_3,\,x_8-x_7} = -\cdot 0696{,}0576, \\
r_{x_2-x_1,\,x_9-x_8} = -\cdot 0725{,}9148, & r_{x_3-x_2,\,x_9-x_8} = -\cdot 0838{,}8890, & r_{x_4-x_3,\,x_9-x_8} = -\cdot 0919{,}5513, \\
r_{x_2-x_1,\,x_{10}-x_9} = -\cdot 0873{,}9443, & r_{x_3-x_2,\,x_{10}-x_9} = -\cdot 1009{,}9575, & r_{x_4-x_3,\,x_{10}-x_9} = -\cdot 1107{,}0675, \\
r_{x_2-x_1,\,x_{11}-x_{10}} = -\cdot 1000{,}0705, & r_{x_3-x_2,\,x_{11}-x_{10}} = -\cdot 1155{,}7130, & r_{x_4-x_3,\,x_{11}-x_{10}} = -\cdot 1266{,}8378.
\end{array}
$$

The numerical results indicate at once the following conclusions:

(*a*) The correlation even between adjacent intervals is not high; it starts positive and rapidly falls, becoming negative, and reaches a negative maximum value when the intervals are farthest apart.

(*b*) A rank-variate and an interval below it are highly, but a rank-variate and an interval above it, i.e. on the side of the tail, are not highly correlated even if adjacent.

(*c*) The change of sign in the interval correlation corresponds to a change of sign in the correlation of the rank-variate with an interval above it.

(*d*) It is clear from the numerical values that

$$r_{x_2-x_1,\, x_4-x_3} = r_{x_2,\, x_2-x_1} \times r_{x_2,\, x_4-x_3}, \quad r_{x_3-x_2,\, x_4-x_3} = r_{x_3,\, x_3-x_2} \times r_{x_3,\, x_4-x_3}, \text{ etc.}$$

This is a general theorem, quite independent of our special parent population, and it may be proved as follows.

If $q < q' < q'' < q''' < q^{\text{iv}}$, let us consider

$$r_{x_{q'}-x_q,\, x_{q^{\text{iv}}}-x_{q'''}} = \frac{\{(x_{q'}-x_q)(x_{q^{\text{iv}}}-x_{q'''})\} - (\overline{x}_{q'}-\overline{x}_q)(\overline{x}_{q^{\text{iv}}}-\overline{x}_{q'''})}{\sigma_{x_{q'}-x_q}\sigma_{x_{q^{\text{iv}}}-x_{q'''}}}$$

$$= \frac{\{x_{q'}x_{q^{\text{iv}}}\} - \overline{x}_{q'}\overline{x}_{q^{\text{iv}}} + \{x_q x_{q'''}\} - \overline{x}_q\overline{x}_{q'''} - \{x_{q'}x_{q'''}\} + \overline{x}_{q'}\overline{x}_{q'''} - \{x_q x_{q^{\text{iv}}}\} + \overline{x}_q\overline{x}_{q^{\text{iv}}}}{\sigma_{x_{q'}-x_q}\sigma_{x_{q^{\text{iv}}}-x_{q'''}}}$$

$$= \frac{r_{x_{q'},\, x_{q^{\text{iv}}}}\sigma_{x_{q'}}\sigma_{x_{q^{\text{iv}}}} + r_{x_q,\, x_{q'''}}\sigma_{x_q}\sigma_{x_{q'''}} - r_{x_{q'},\, x_{q'''}}\sigma_{x_{q'}}\sigma_{x_{q'''}} - r_{x_q,\, x_{q^{\text{iv}}}}\sigma_{x_q}\sigma_{x_{q^{\text{iv}}}}}{\sigma_{x_{q'}-x_q}\sigma_{x_{q^{\text{iv}}}-x_{q'''}}}.$$

But

$$r_{x_{q'},\, x_{q^{\text{iv}}}} = r_{x_{q'},\, x_{q''}} r_{x_{q''},\, x_{q^{\text{iv}}}}, \quad r_{x_{q'},\, x_{q'''}} = r_{x_{q'},\, x_{q''}} r_{x_{q''},\, x_{q'''}},$$

$$r_{x_q,\, x_{q'''}} = r_{x_q,\, x_{q''}} r_{x_{q''},\, x_{q'''}}, \quad r_{x_q,\, x_{q^{\text{iv}}}} = r_{x_q,\, x_{q''}} r_{x_{q''},\, x_{q^{\text{iv}}}},$$

by formula (ciii). Hence

$$r_{x_{q'}-x_q,\, x_{q^{\text{iv}}}-x_{q'''}} = \frac{(\sigma_{x_{q'}} r_{x_{q'},\, x_{q''}} - \sigma_{x_q} r_{x_q,\, x_{q''}})(\sigma_{x_{q^{\text{iv}}}} r_{x_{q''},\, x_{q^{\text{iv}}}} - \sigma_{x_{q'''}} r_{x_{q''},\, x_{q'''}})}{\sigma_{x_{q'}-x_q}\sigma_{x_{q^{\text{iv}}}-x_{q'''}}} \quad \text{............(cxxi)}$$

$$= \frac{(\sigma_{x_{q''}}\sigma_{x_{q'}} r_{x_{q'},\, x_{q''}} - \sigma_{x_{q''}}\sigma_{x_q} r_{x_q,\, x_{q''}})(\sigma_{x_{q''}}\sigma_{x_{q^{\text{iv}}}} r_{x_{q''},\, x_{q^{\text{iv}}}} - \sigma_{x_{q''}}\sigma_{x_{q'''}} r_{x_{q''},\, x_{q'''}})}{\sigma_{x_{q''}}\sigma_{x_q - x_q} \times \sigma_{x_{q''}}\sigma_{x_{q''}-x_{q'''}}}$$

$$= \frac{\{x_{q''}x_{q'}\} - \overline{x}_{q'}\overline{x}_{q''} - \{x_{q''}x_q\} + \overline{x}_q\overline{x}_{q''}}{\sigma_{x_{q''}}\sigma_{x_{q'}-x_q}} \times \frac{\{x_{q^{\text{iv}}}x_{q''}\} - \overline{x}_{q^{\text{iv}}}\overline{x}_{q''} - \{x_{q'''}x_{q''}\} + \overline{x}_{q'''}\overline{x}_{q''}}{\sigma_{x_{q''}}\sigma_{x_{q^{\text{iv}}}-x_{q'''}}}$$

$$= \frac{\{x_{q''}(x_{q'}-x_q)\} - \overline{x}_{q''}(\overline{x}_{q'}-\overline{x}_q)}{\sigma_{x_{q''}}\sigma_{x_{q'}-x_q}} \times \frac{\{x_{q''}(x_{q^{\text{iv}}}-x_{q'''})\} - \overline{x}_{q''}(\overline{x}_{q^{\text{iv}}}-\overline{x}_{q'''})}{\sigma_{x_{q''}}(\sigma_{x_{q^{\text{iv}}}-x_{q'''}})}$$

$$= r_{x_{q''},\, x_{q'-q}} \times r_{x_{q''},\, x_{q^{\text{iv}}}-x_{q'''}} \quad \text{......................................(cxxii).}$$

We thus find that the partial correlation

$$r_{x_{q'-q},\, x_{q^{\text{iv}}}-x_{q'''}\cdot x_{q''}} = 0,$$

or, if we fix the variate of any rank q'', then the correlation between any two inter-rank intervals, one on either side of q'', is zero. This clearly follows from the general principle, that fixing the variate of any rank is equivalent to dividing the parental population into two portions and then taking at random $q''-1$ individuals from the first portion and $n-q''$ from the second portion; the distribution of these two sets of individuals must be completely independent.

Formula (cxxii) is a convenient form for testing the accuracy with which any interrank distance correlation has been determined. If we remember that any *interrank distance variance can be written as*

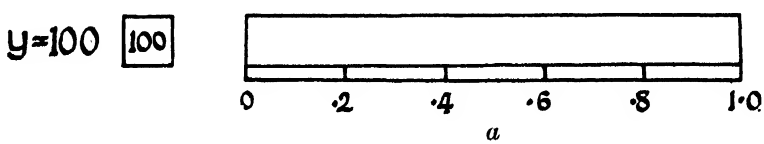
$$\left.\begin{aligned} \sigma^2_{x_{q'}-x_q} &= \sigma^2_{x_q} + \sigma^2_{x_{q'}} - 2r_{x_q x_{q'}}\sigma_{x_q}\sigma_{x_{q'}} \\ \text{and that}\qquad r_{x_q,\, x_{q''}-x_{q'}} &= \frac{\sigma_{x_{q''}} r_{x_q,\, x_{q''}} - \sigma_{x_{q'}} r_{x_q,\, x_{q'}}}{x_{q''}-x_{q'}} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(\text{cxxiii}),$$

it will be clear that the whole system of correlations between the rank-variates and the interrank distances can be reduced by aid of formulae (ciii), (cxxii) and (cxxiii) to the calculation of the σ_{x_q}'s and the $r_{x_q x_{q+1}}$'s. The work can thus be very much simplified. Ultimately it depends on a knowledge of $\{x_q\}$, $\{x_q^2\}$, $\{x_q x_{q+1}\}$* throughout the whole n ranks of the sample. And this is true whatever be the character of the parental population, supposed continuous.

Illustration III.

Rectangular Parent Populations. Curve (iii). The change from the asymptotic curves is indicated in Fig. III. $y = N/b$, $\sigma = {\cdot}288{,}675b$, $\beta_1 = 0$, $\beta_2 = 1{\cdot}8$. The theory has been given in Part I of this paper.

FIG: III



Additional formulae which may be of service are

$$\sigma_{x_{q'}-x_q} = \sqrt{\frac{(q'-q)(n+1-q'+q)}{n+2}} \cdot \frac{b}{n+1},$$

which, if $q' = q+1$, gives

$$\sigma_{x_{q+1}-x_q} = \sqrt{\frac{n}{n+2}}\,\frac{b}{n+1},$$

a value independent of q, and if $q < q' < q'' < q'''$,

$$r_{x_{q'}-x_q,\, x_{q'''}-x_{q''}} = -\sqrt{\frac{(q'-q)(q'''-q'')}{(n+1-q'+q)(n+1-q^{\text{iv}}-q'')}},$$

or if $q' = q+1$, $q''' = q''+1$,

$$r_{x_{q+1}-x_q,\, x_{q''+1}-x_{q''}} = -\frac{1}{n};$$

a quantity independent of both q and q'', or all rank-intervals have the same negative correlation.

* It is not indeed needful in the cases under consideration to find $\{x_q x_{q'}\}$, for $r_{x_q x_{q'}} = \dfrac{\sigma_{x_{q'}}}{\bar{x}_{q'}} \Big/ \dfrac{\sigma_{x_q}}{\bar{x}_q}$ by (cvii), a result which shows us that the relative variability always must decrease with increasing rank.

In the case of samples of 11,

$$\sigma_{x_{q+1}-x_q} = \cdot 0766{,}5551{,}76b,$$
$$r_{x_{q+1}-x_q,\, x_{q''+1}-x_{q''}} = -\cdot 0909{,}0909{,}09.$$

Comparing the first of these results with those on p. 256, a being the total range of curve and b of rectangle, we see that the result is obtained by the middle interval $(x_6 - x_5)$ almost retaining its variability while those below creep up to its value, and those above fall down to it.

Comparing the second result with those on p. 258, we note that the negative correlation has spread right through the whole system, and then the values have been equalised as we pass from the highest to the lowest at a value less than the highest.

Further, if $q'' > q' > q$,

$$r_{x_{q''},\, x_{q'}-x_q} = \sqrt{\frac{(q'-q)(n+1-q'')}{q''(n+1-q'+q)}};$$

if $q'' = q'$,

$$r_{x_{q'},\, x_{q'}-x_q} = \sqrt{\frac{(q'-q)(n+1-q')}{q'(n+1-q'+q)}},$$

if, in addition, $q' = q+1$,

$$r_{x_{q+1},\, x_{q+1}-x_q} = \sqrt{\frac{n-q}{n(q+1)}}.$$

Again, $$r_{x_q,\, x_{q'}-x_q} = -\sqrt{\frac{q(q'-q)}{(n+1-q)(n+1-q'+q)}};$$

if $q' = q+1$,

$$r_{x_q,\, x_{q+1}-x_q} = -\sqrt{\frac{q}{n(n+1-q)}}.$$

Hence we find for $n = 11$,

$$\begin{array}{ll}
r_{x_2,\, x_2-x_1} = \cdot 6741{,}9986, & r_{x_1,\, x_2-x_1} = -\cdot 1348{,}3997, \\
r_{x_3,\, x_2-x_1} = \cdot 5222{,}3297, & r_{x_2,\, x_3-x_2} = -\cdot 1348{,}3997, \\
r_{x_3,\, x_3-x_2} = \cdot 5222{,}3297, & r_{x_2,\, x_4-x_3} = -\cdot 1740{,}7766, \\
r_{x_4,\, x_2-x_1} = \cdot 4264{,}0143, & r_{x_3,\, x_5-x_4} = -\cdot 1740{,}7766, \\
r_{x_4,\, x_3-x_2} = \cdot 4264{,}0143, & r_{x_4,\, x_5-x_4} = -\cdot 2132{,}0072, \\
r_{x_4,\, x_4-x_3} = \cdot 4264{,}0143, & r_{x_4,\, x_6-x_5} = -\cdot 2132{,}0072.
\end{array}$$

We see at once from these results that the correlation of the variate of any rank with any rank-interval below it is constant. This follows at once, if we put $q' = q+1$, when we find

$$r_{x_{q''},\, x_{q+1}-x_q} = \sqrt{\frac{n+1-q''}{nq''}},$$

which is independent of q.

A similar property holds for the correlation of the variate of any rank with any interval *above* it; for

$$r_{x_q,\, x_{q''}-x_{q'}} = -\sqrt{\frac{q(q''-q')}{(n+1-q)(n+1-q''+q')}}.$$

Hence, if $q'' = q' + 1$,

$$r_{x_q, x_{q'+1} - x_{q'}} = -\sqrt{\frac{q}{n(n+1-q)}},$$

which is independent of q'.

This relation is also obvious in the above table of the correlations.

Clearly, the correlation of variate of qth rank with any interval below it × correlation of variate of qth rank with any other interval above it $= -\frac{1}{n} =$ correlation of the first interval with the second interval, e.g.

$$r_{x_2, x_2 - x_1} \times r_{x_2, x_3 - x_2} = -\tfrac{1}{11} = r_{x_2 - x_1, x_3 - x_2},$$

$$r_{x_4, x_4 - x_3} \times r_{x_4, x_6 - x_5} = -\tfrac{1}{11} = r_{x_4 - x_3, x_6 - x_5},$$

etc.

This is a special case of the theorem already stated (see p. 259) emphasised by the fact that here the correlation of any pair of intervals $= -\frac{1}{n} = -\frac{1}{11}$.

Turning from the correlations to the actual distribution of rank-variates, we have the mean variates of the successive ranks distributed at $\frac{1}{12}b$ apart in the case of samples of 11 starting at distance $\frac{1}{12}b$ from the start of the range. We have the following symmetrical system of standard deviations:

$$\sigma_{x_1} = \cdot 0766{,}5551{,}76b = \sigma_{x_{11}},$$
$$\sigma_{x_2} = \cdot 1033{,}6227{,}88b = \sigma_{x_{10}},$$
$$\sigma_{x_3} = \cdot 1201{,}2813{,}67b = \sigma_{x_9},$$
$$\sigma_{x_4} = \cdot 1307{,}4409{,}01b = \sigma_{x_8},$$
$$\sigma_{x_5} = \cdot 1367{,}3544{,}24b = \sigma_{x_7},$$
$$\sigma_{x_6} = \cdot 1386{,}7504{,}91b.$$

The variation rises from a minimum at the first and last ranks, and afterwards more slowly reaches a maximum at the median. Since x_5, x_6 and x_7 have the highest variation, it might be expected that their intervals would have, but this high variation is accompanied by a high correlation, which causes the variation of all intervals to be the same.

This can be shewn on the rank-variate correlations. Here we have*

$$r_{x_1, x_2} = \cdot 6741{,}9986 = r_{x_{10}, x_{11}},$$
$$r_{x_2, x_3} = \cdot 7745{,}9667 = r_{x_9, x_{10}},$$
$$r_{x_3, x_4} = \cdot 8164{,}9658 = r_{x_8, x_9},$$
$$r_{x_4, x_5} = \cdot 8366{,}6003 = r_{x_7, x_8},$$
$$r_{x_5, x_6} = \cdot 8451{,}5425 = r_{x_6, x_7}.$$

The rank-variates and rank-intervals when we sample from a rectangular population being so easy to analyse, it is somewhat sad that such parental distributions are so rare as to be of little practical importance.

* $= \sqrt{\frac{q}{q'}\frac{n+1-q'}{n+1-q}}$, see *Biometrika*, Vol. XXIII, p. 391 (xxvii).

Illustration IV. Type IX.

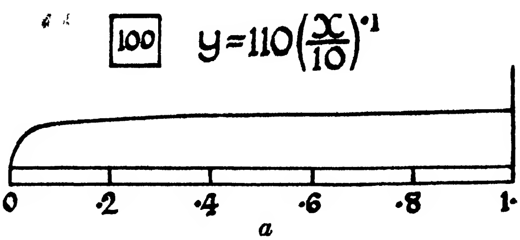
$$\text{Curve (iv)}: \ y = 110\left(\frac{x}{a}\right)^{0\cdot 1}$$

Range from $x = 0$ to $x = a$.

$$\sigma = a \times \cdot 2836{,}5876, \quad \beta_1 = \cdot 006{,}706, \quad \beta_2 = 1\cdot 831{,}616.$$

This curve is drawn in Fig. IV, and indicates the rectangle changing to the sloping straight line, there is here no asymptote, but still a finite range, and finite ordinate at a. As the power increases the ordinate at a becomes greater, and the

FIG: IV



curvature changes from concave to convex at the line point, see Figs. VI, VII, and VIII, below. Ultimately the range becomes infinite, while the ordinate remains finite at the Exponential Point.

These curves fall under Type IX with the formulae for samples of 11:

$$\bar{x}_q = \frac{\Gamma(12)}{\Gamma(q)} \frac{\Gamma(q+\frac{10}{11})}{\Gamma(12+\frac{10}{11})},$$

$$\{\bar{x}_q^2\} = \frac{\Gamma(12)}{\Gamma(q)} \frac{\Gamma(q+1+\frac{9}{11})}{\Gamma(13+\frac{9}{11})},$$

$$\{x_q x_{q'}\} = \frac{\Gamma(12)}{\Gamma(13+\frac{9}{11})} \frac{\Gamma(q'+1+\frac{9}{11})}{\Gamma(q'+\frac{10}{11})} \frac{\Gamma(q+\frac{10}{11})}{\Gamma(q)}.$$

From these results we obtain

	For Curve (iv)	For Rectangle
$\bar{x}_1$ =	·1011,3289,73a,	·0833,3333,33a,
$\bar{x}_2$ =	·1930,7189,49a,	·1666,6666,67a,
$\bar{x}_3$ =	·2808,3184,72a,	·2500,0000,00a,
$\bar{x}_4$ =	·3659,3240,69a,	·3333,3333,33a,
$\bar{x}_5$ =	·4490,9886,30a,	·4166,6666,67a,
$\bar{x}_6$ =	·5307,5320,17a,	·5000,0000,00a,
$\bar{x}_7$ =	·6111,7035,35a,	·5833,3333,33a,
$\bar{x}_8$ =	·6905,4312,67a,	·6666,6666,67a,
$\bar{x}_9$ =	·7690,1393,66a,	·7500,0000,00a,
$\bar{x}_{10}$ =	·8466,9211,20a,	·8333,3333,33a,
$\bar{x}_{11}$ =	·9236,6412,21a,	·9166,6666,67a.

It will be seen that there has been a general shift of the mean variates of the ranks toward the terminal of the range as measured from the origin, the shift diminishing though not very rapidly for the higher ranks.

If we compare the corresponding variabilities

	For Curve (iv)	For Rectangle
σ_{x_1}	$=\cdot 0852,5455,75a,$	$\cdot 0766,5551,76a,$
σ_{x_2}	$=\cdot 1096,8465,31a,$	$\cdot 1033,6227,88a,$
σ_{x_3}	$=\cdot 1235,5539,22a,$	$\cdot 1201,2813,67a,$
σ_{x_4}	$=\cdot 1314,3786,15a,$	$\cdot 1307,4409,01a,$
σ_{x_5}	$=\cdot 1349,5161,74a,$	$\cdot 1367,3544,24a,$
σ_{x_6}	$=\cdot 1347,8513,03a,$	$\cdot 1386,7504,91a,$
σ_{x_7}	$=\cdot 1311,4557,68a,$	$\cdot 1367,3544,24a,$
σ_{x_8}	$=\cdot 1239,9249,66a,$	$\cdot 1307,4409,01a,$
σ_{x_9}	$=\cdot 1127,4126,88a,$	$\cdot 1201,2813,67a,$
$\sigma_{x_{10}}$	$=\cdot 0961,4827,86a,$	$\cdot 1033,6227,88a,$
$\sigma_{x_{11}}$	$=\cdot 0707,1504,73a,$	$\cdot 0766,5551,76a,$

we see that the effect of lessening the frequency at the origin terminal of the range is to increase the variability of the first four ranks, and reduce that of the remaining seven ranks. Generally in skew limited range curves the variates of the ranks towards the end with lesser frequency have higher variabilities than those towards that end with greater frequency.

We next note that while the rank variates have changed considerably in mean and variability, their correlation system has remained nearly symmetrical and very close to that for the rectangle. The reader will find this to be so by comparing the following values with those on p. 262:

$$r_{x_1, x_2} = \cdot 6739,0962, \quad r_{x_{10}, x_{11}} = \cdot 6741,8914,$$
$$r_{x_2, x_3} = \cdot 7744,4136, \quad r_{x_9, x_{10}} = \cdot 7745,8181,$$
$$r_{x_3, x_4} = \cdot 8164,0250, \quad r_{x_8, x_9} = \cdot 8164,7728,$$
$$r_{x_4, x_5} = \cdot 8365,9743, \quad r_{x_7, x_8} = \cdot 8367,8372,$$
$$x_{x_5, x_6} = \cdot 8451,0996, \quad x_{x_6, x_7} = \cdot 8452,7165.$$

We now turn to the rank-intervals and their standard deviations. Here

$$\bar{x}_2 - \bar{x}_1 = \cdot 0919,3899,76a, \quad \sigma_{x_2 - x_1} = \cdot 0818,2571,71a,$$
$$\bar{x}_3 - \bar{x}_2 = \cdot 0877,5995,23a, \quad \sigma_{x_3 - x_2} = \cdot 0794,1031,13a,$$
$$\bar{x}_4 - \bar{x}_3 = \cdot 0851,0055,97a, \quad \sigma_{x_4 - x_3} = \cdot 0776,2300,67a,$$
$$\bar{x}_5 - \bar{x}_4 = \cdot 0831,6645,61a, \quad \sigma_{x_5 - x_4} = \cdot 0762,1768,17a,$$
$$\bar{x}_6 - \bar{x}_5 = \cdot 0816,5433,87a, \quad \sigma_{x_6 - x_5} = \cdot 0750,6506,20a,$$
$$\bar{x}_7 - \bar{x}_6 = \cdot 0804,1715,18a, \quad \sigma_{x_7 - x_6} = \cdot 0740,4964,31a,$$
$$\bar{x}_8 - \bar{x}_7 = \cdot 0793,7277,32a, \quad \sigma_{x_8 - x_7} = \cdot 0732,0729,69a,$$
$$\bar{x}_9 - \bar{x}_8 = \cdot 0784,7080,99a, \quad \sigma_{x_9 - x_8} = \cdot 0725,0892,79a,$$
$$\bar{x}_{10} - \bar{x}_9 = \cdot 0776,7817,54a, \quad \sigma_{x_{10} - x_9} = \cdot 0718,4942,03a,$$
$$\bar{x}_{11} - \bar{x}_{10} = \cdot 0769,7201,01a, \quad \sigma_{x_{11} - x_{10}} = \cdot 0712,5524,56a.$$

In the case of the rectangle the mean interrank intervals are all

$$= \frac{a}{12} = \cdot 0833{,}3333{,}33a,$$

or the effect of reducing the frequency at the start of the range is to increase the rank-interval here, and this increase for samples of 11 holds for the first three intervals, after which the intervals are reduced. Again, the variability of the intervals in the rectangle is constant $= \cdot 0766{,}5551{,}76a$; hence the lessened terminal frequency increases the variability of the intervals up to the third, after which it becomes slowly less.

We may next consider the correlations between adjacent intervals; we have

$$r_{x_2-x_1,\, x_3-x_2} = -\cdot 1125{,}2098, \quad r_{x_{10}-x_9,\, x_{11}-x_{10}} = -\cdot 0835{,}1692,$$
$$r_{x_3-x_2,\, x_4-x_3} = -\cdot 1017{,}8331, \quad r_{x_9-x_8,\, x_{10}-x_9} = -\cdot 0845{,}0094,$$
$$r_{x_4-x_3,\, x_5-x_4} = -\cdot 0957{,}7702, \quad r_{x_8-x_7,\, x_9-x_8} = -\cdot 0857{,}6100,$$
$$r_{x_5-x_4,\, x_6-x_5} = -\cdot 0919{,}1912, \quad r_{x_7-x_6,\, x_8-x_7} = -\cdot 0873{,}3809,$$
$$r_{x_6-x_5,\, x_7-x_6} = -\cdot 0892{,}7658.$$

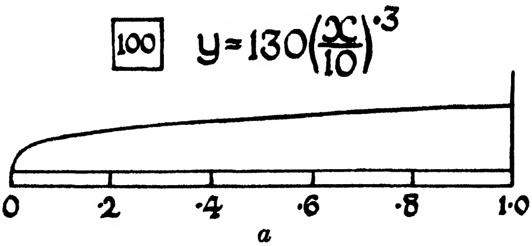
All these correlations are for the rectangle $-\cdot 0909{,}0909$. Thus the effect of lessening the frequency at the start of the range is to increase the negative correlation between adjacent interrank intervals at that end, and decrease it at the other terminal. $m =$ the power, 0·1 is so small, however, that there are not very great differences from the rectangular value. We have none of the positive correlations which marked the relation of interrank intervals for curves on the biquadratic above the rectangular point.

Illustration V. Type IX.

$$\text{Curve (v)}: \; y = 130\left(\frac{x}{a}\right)^{0\cdot 3}. \quad \text{Range } x = 0 \text{ to } a.$$

$$\sigma = \cdot 2728{,}8953a, \quad \beta_1 = \cdot 049{,}424, \quad \beta_2 = 1\cdot 928{,}072.$$

FIG: V



For this curve we find

$$\bar{x}_1 = \cdot 1375{,}6917{,}70a, \quad \sigma_{x_1} = \cdot 0998{,}0335{,}15a,$$
$$\bar{x}_2 = \cdot 2433{,}9162{,}08a, \quad \sigma_{x_2} = \cdot 1186{,}8650{,}72a,$$
$$\bar{x}_3 = \cdot 3370{,}0378{,}26a, \quad \sigma_{x_3} = \cdot 1271{,}1809{,}12a,$$
$$\bar{x}_4 = \cdot 4234{,}1500{,}89a, \quad \sigma_{x_4} = \cdot 1303{,}0265{,}11a,$$
$$\bar{x}_5 = \cdot 5048{,}4097{,}21a, \quad \sigma_{x_5} = \cdot 1299{,}1879{,}68a,$$

$$
\begin{array}{llll}
\bar{x}_6 & = \cdot 5825{,}0881{,}40a, & \sigma_{x_6} & = \cdot 1266{,}4953{,}94a, \\
\bar{x}_7 & = \cdot 6571{,}8943{,}12a, & \sigma_{x_7} & = \cdot 1207{,}3010{,}76a, \\
\bar{x}_8 & = \cdot 7294{,}0805{,}00a, & \sigma_{x_8} & = \cdot 1120{,}8788{,}29a, \\
\bar{x}_9 & = \cdot 7995{,}4343{,}94a, & \sigma_{x_9} & = \cdot 1003{,}0361{,}70a, \\
\bar{x}_{10} & = \cdot 8678{,}8048{,}55a, & \sigma_{x_{10}} & = \cdot 0843{,}2473{,}69a, \\
\bar{x}_{11} & = \cdot 9346{,}4052{,}29a & \sigma_{x_{11}} & = \cdot 0612{,}1851{,}39a.
\end{array}
$$

Further,

$$
\begin{array}{llll}
\bar{x}_2 - \bar{x}_1 & = \cdot 1058{,}2244{,}38a, & \sigma_{x_2-x_1} & = \cdot 0901{,}2957{,}95a, \\
\bar{x}_3 - \bar{x}_2 & = \cdot 0936{,}1216{,}18a, & \sigma_{x_3-x_2} & = \cdot 0830{,}9455{,}49a, \\
\bar{x}_4 - \bar{x}_3 & = \cdot 0864{,}1122{,}63a, & \sigma_{x_4-x_3} & = \cdot 0781{,}6868{,}42a, \\
\bar{x}_5 - \bar{x}_4 & = \cdot 0814{,}2596{,}32a, & \sigma_{x_5-x_4} & = \cdot 0744{,}6268{,}59a, \\
\bar{x}_6 - \bar{x}_5 & = \cdot 0776{,}6784{,}19a, & \sigma_{x_6-x_5} & = \cdot 0715{,}1999{,}34a, \\
\bar{x}_7 - \bar{x}_6 & = \cdot 0746{,}8061{,}72a, & \sigma_{x_7-x_6} & = \cdot 0691{,}1598{,}52a, \\
\bar{x}_8 - \bar{x}_7 & = \cdot 0722{,}1861{,}88a, & \sigma_{x_8-x_7} & = \cdot 0670{,}8148{,}41a, \\
\bar{x}_9 - \bar{x}_8 & = \cdot 0701{,}3538{,}94a, & \sigma_{x_9-x_8} & = \cdot 0653{,}2942{,}15a, \\
\bar{x}_{10} - \bar{x}_9 & = \cdot 0683{,}3704{,}61a, & \sigma_{x_{10}-x_9} & = \cdot 0637{,}9610{,}84a, \\
\bar{x}_{11} - \bar{x}_{10} & = \cdot 0667{,}6003{,}73a, & \sigma_{x_{11}-x_{10}} & = \cdot 0624{,}3669{,}88a.
\end{array}
$$

The correlations between adjacent segments are

$$
\begin{array}{llll}
r_{x_2-x_1,\, x_3-x_2} & = -\cdot 1389{,}2856, & r_{x_{11}-x_{10},\, x_{10}-x_9} & = -\cdot 0717{,}7055, \\
r_{x_3-x_2,\, x_4-x_3} & = -\cdot 1131{,}2660, & r_{x_{10}-x_9,\, x_9-x_8} & = -\cdot 0738{,}7633, \\
r_{x_4-x_3,\, x_5-x_4} & = -\cdot 0989{,}5414, & r_{x_9-x_8,\, x_8-x_7} & = -\cdot 0764{,}8528, \\
r_{x_5-x_4,\, x_6-x_5} & = -\cdot 0901{,}7566, & r_{x_8-x_7,\, x_7-x_6} & = -\cdot 0798{,}0314,
\end{array}
$$

$$r_{x_6-x_5,\, x_7-x_6} = -\cdot 0841{,}7487.$$

These values enable us to build up other correlations, as we have already indicated. For example,

$$r_{x_2-x_1,\, x_4-x_3\,.\,(x_3-x_2)} = 0 = r_{x_2-x_1,\, x_4-x_3} - r_{x_2-x_1,\, x_3-x_2} \times r_{x_4-x_3,\, x_3-x_2},$$

and thus

$$
\begin{aligned}
r_{x_2-x_1,\, x_4-x_3} &= -\cdot 1389{,}2856 \times (-\cdot 1131{,}2660) \\
&= \cdot 0157{,}1652.
\end{aligned}
$$

Again,

$$r_{x_2-x_1,\, x_5-x_4\,.\,(x_4-x_3)} = 0,$$

or

$$
\begin{aligned}
r_{x_2-x_1,\, x_5-x_4} &= r_{x_2-x_1,\, x_4-x_3} \times r_{x_5-x_4,\, x_4-x_3} \\
&= \cdot 0157{,}1652 \times (-\cdot 0989{,}5414) \\
&= -\cdot 0015{,}5521,
\end{aligned}
$$

which sufficiently indicate the alternating in sign of the correlation of successive intervals with a given interval, and the rapid reduction in intensity of the association.

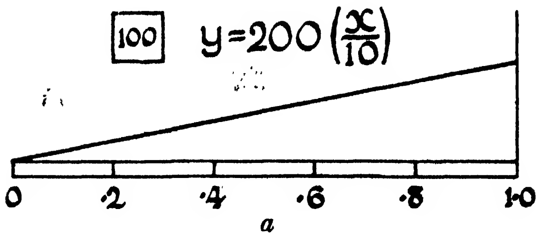
Generally we note that the whole system of rank-variates is shifting away from the curtailed end of the range, that the intervals are becoming larger at this end and smaller at the other, while the variability of both rank-variates and intervals is becoming larger at the same end and smaller at the other also.

Illustration VI. Type IX.

$$\text{The Line-Curve (vi): } y = 200\left(\frac{x}{a}\right).$$

$$\sigma = \cdot 235{,}7023a, \quad \beta_1 = \cdot 32, \quad \beta_2 = 2\cdot 4.$$

FIG: VI



$\bar{x}_1 = \cdot 2585{,}0974{,}08a,$	$\sigma_{x_1} = \cdot 1284{,}7586{,}25a,$
$\bar{x}_2 = \cdot 3877{,}6461{,}12a,$	$\sigma_{x_2} = \cdot 1276{,}9210{,}22a,$
$\bar{x}_3 = \cdot 4847{,}0576{,}40a,$	$\sigma_{x_3} = \cdot 1227{,}2050{,}50a,$
$\bar{x}_4 = \cdot 5654{,}9005{,}80a,$	$\sigma_{x_4} = \cdot 1164{,}2305{,}45a,$
$\bar{x}_5 = \cdot 6361{,}7631{,}53a,$	$\sigma_{x_5} = \cdot 1092{,}9941{,}68a,$
$\bar{x}_6 = \cdot 6997{,}9394{,}68a,$	$\sigma_{x_6} = \cdot 1014{,}3190{,}83a,$
$\bar{x}_7 = 7581{,}1010{,}90a,$	$\sigma_{x_7} = \cdot 0927{,}5395{,}38a,$
$\bar{x}_8 = \cdot 8122{,}6083{,}11a,$	$\sigma_{x_8} = \cdot 0830{,}6027{,}29a,$
$\bar{x}_9 = \cdot 8630{,}2713{,}31a,$	$\sigma_{x_9} = \cdot 0720{,}0116{,}40a,$
$\bar{x}_{10} = \cdot 9109{,}7308{,}49a,$	$\sigma_{x_{10}} = \cdot 0588{,}3342{,}52a,$
$\bar{x}_{11} = \cdot 9565{,}2173{,}91a,$	$\sigma_{x_{11}} = \cdot 0416{,}2726{,}63a.$

$\bar{x}_2 - \bar{x}_1 = \cdot 1292{,}5487{,}04a,$	$\sigma_{x_2-x_1} = \cdot 1052{,}1861{,}17a,$
$\bar{x}_3 - \bar{x}_2 = \cdot 0969{,}4115{,}28a,$	$\sigma_{x_3-x_2} = \cdot 0852{,}5889{,}73a,$
$\bar{x}_4 - \bar{x}_3 = \cdot 0807{,}8429{,}40a,$	$\sigma_{x_4-x_3} = \cdot 0733{,}3934{,}65a,$
$\bar{x}_5 - \bar{x}_4 = \cdot 0706{,}8625{,}73a,$	$\sigma_{x_5-x_4} = \cdot 0652{,}8945{,}01a,$
$\bar{x}_6 - \bar{x}_5 = \cdot 0636{,}1763{,}15a,$	$\sigma_{x_6-x_5} = \cdot 0594{,}0163{,}75a,$
$\bar{x}_7 - \bar{x}_6 = \cdot 0583{,}1616{,}22a,$	$\sigma_{x_7-x_6} = \cdot 0548{,}5874{,}26a,$
$\bar{x}_8 - \bar{x}_7 = \cdot 0541{,}5072{,}21a,$	$\sigma_{x_8-x_7} = \cdot 0512{,}1772{,}01a,$
$\bar{x}_9 - \bar{x}_8 = \cdot 0507{,}6630{,}19a,$	$\sigma_{x_9-x_8} = \cdot 0482{,}1559{,}27a,$
$\bar{x}_{10} - \bar{x}_9 = \cdot 0479{,}4595{,}18a,$	$\sigma_{x_{10}-x_9} = \cdot 0456{,}8534{,}36a,$
$\bar{x}_{11} - \bar{x}_{10} = \cdot 0455{,}4865{,}42a,$	$\sigma_{x_{11}-x_{10}} = \cdot 0435{,}1521{,}65a.$

The straight line takes its place in the general series of these curves with no simplification in the results, and only a slight lessening of the labour of determining them.

The correlations between successive interrank intervals are

$$r_{x_2-x_1,\,x_3-x_2} = -\cdot 1581{,}8003, \quad r_{x_{11}-x_{10},\,x_{10}-x_9} = -\cdot 0479{,}4693,$$
$$r_{x_3-x_2,\,x_4-x_3} = -\cdot 1101{,}0458, \quad r_{x_{10}-x_9,\,x_9-x_8} = -\cdot 0508{,}7161,$$
$$r_{x_4-x_3,\,x_5-x_4} = -\cdot 0875{,}7500, \quad r_{x_9-x_8,\,x_8-x_7} = -\cdot 0545{,}2651,$$
$$r_{x_5-x_4,\,x_6-x_5} = -\cdot 0742{,}9493, \quad r_{x_8-x_7,\,x_7-x_6} = -.0592{,}2374,$$
$$r_{x_6-x_5,\,x_7-x_6} = -\cdot 0654{,}9627.$$

This series of correlations still further emphasises the increased negative coefficients where the frequency is cut down, and the reduced coefficients at the mode.

If we find the correlation of a rank-variate with the interval on either of that rank, we have the usual rule obeyed, i.e.

$$r_{x_{q+1}-x_q,\,x_q-x_{q-1}} = r_{x_q,\,x_q-x_{q-1}} \times r_{x_q,\,x_{q+1}-x_q}.$$

For example, $$r_{x_2,\,x_2-x_1} = (\{x_2^2\} - \{x_2x_1\} - \bar{x}_2(\bar{x}_2 - \bar{x}_1))/\sigma_{x_2}\sigma_{x_2-x_1}$$
$$= \cdot 4045{,}29516,$$
$$r_{x_2,\,x_3-x_2} = (\{x_2x_3\} - \{x_2^2\} - \bar{x}_2(\bar{x}_3 - \bar{x}_2))/\sigma_{x_2}\sigma_{x_3-x_2}$$
$$= -\cdot 3910{,}22224,$$

and accordingly their product

$$= -\cdot 1581{,}8003 = r_{x_2-x_1,\,x_3-x_2}.$$

The correlations between adjacent rank-variates are high and positive. Thus

$$r_{x_1,\,x_2} = \cdot 6625{,}9970, \quad r_{x_{10},\,x_{11}} = \cdot 6738{,}5188, \text{ and } r_{x_5,\,x_6} = \cdot 8436{,}5341,$$

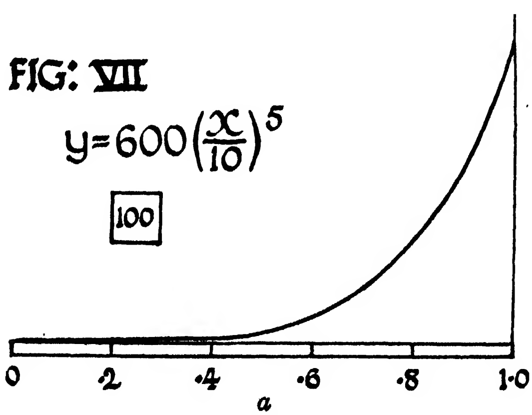
which show comparatively little change from the values for $y = y_0\left(\frac{x}{a}\right)^1$: see p. 264.

Illustration VII. Type IX.

$$\text{Curve (vii)}: \; y = 600\left(\frac{x}{a}\right)^5.$$

$$\text{Range } x = 0 \text{ to } a.$$

$$\sigma = \cdot 1237{,}1791a, \quad \beta_1 = 1\cdot 646{,}091, \quad \beta_2 = 4\cdot 622{,}222.$$



$$\begin{aligned}
\bar{x}_1 &= \cdot 6167{,}2164{,}29a, & \sigma_{x_1} &= \cdot 1155{,}1344{,}27a,\\
\bar{x}_2 &= \cdot 7195{,}0858{,}34a, & \sigma_{x_2} &= \cdot 0789{,}0491{,}38a,\\
\bar{x}_3 &= \cdot 7794{,}6763{,}20a, & \sigma_{x_3} &= \cdot 0695{,}3580{,}97a,\\
\bar{x}_4 &= \cdot 8227{,}7138{,}93a, & \sigma_{x_4} &= \cdot 0591{,}3764{,}57a,\\
\bar{x}_5 &= \cdot 8570{,}5353{,}06a, & \sigma_{x_5} &= \cdot 0511{,}2186{,}64a,\\
\bar{x}_6 &= \cdot 8856{,}2198{,}16a, & \sigma_{x_6} &= \cdot 0444{,}0178{,}54a,\\
\bar{x}_7 &= \cdot 9102{,}2259{,}22a, & \sigma_{x_7} &= \cdot 0384{,}1688{,}84a,\\
\bar{x}_8 &= \cdot 9318{,}9455{,}87a, & \sigma_{x_8} &= \cdot 0328{,}0948{,}14a,\\
\bar{x}_9 &= \cdot 9513{,}0902{,}86a, & \sigma_{x_9} &= \cdot 0272{,}8356{,}92a,\\
\bar{x}_{10} &= \cdot 9689{,}2586{,}25a, & \sigma_{x_{10}} &= \cdot 0214{,}8511{,}95a,\\
\bar{x}_{11} &= \cdot 9850{,}7462{,}69a, & \sigma_{x_{11}} &= \cdot 0147{,}0424{,}20a.
\end{aligned}$$

The variability of the rank-variate now continuously falls.

$$\begin{aligned}
\bar{x}_2 - \bar{x}_1 &= \cdot 1027{,}8694{,}05a, & \sigma_{x_2-x_1} &= \cdot 0904{,}5398{,}74a,\\
\bar{x}_3 - \bar{x}_2 &= \cdot 0599{,}5904{,}86a, & \sigma_{x_3-x_2} &= \cdot 0559{,}8768{,}70a,\\
\bar{x}_4 - \bar{x}_3 &= \cdot 0433{,}0375{,}73a, & \sigma_{x_4-x_3} &= \cdot 0413{,}0473{,}65a,\\
\bar{x}_5 - \bar{x}_4 &= \cdot 0342{,}8214{,}12a, & \sigma_{x_5-x_4} &= \cdot 0330{,}5890{,}82a,\\
\bar{x}_6 - \bar{x}_5 &= \cdot 0285{,}6845{,}10a, & \sigma_{x_6-x_5} &= \cdot 0277{,}3303{,}91a,\\
\bar{x}_7 - \bar{x}_6 &= \cdot 0246{,}0061{,}06a, & \sigma_{x_7-x_6} &= \cdot 0239{,}8827{,}45a,\\
\bar{x}_8 - \bar{x}_7 &= \cdot 0216{,}7196{,}65a, & \sigma_{x_8-x_7} &= \cdot 0212{,}0056{,}16a,\\
\bar{x}_9 - \bar{x}_8 &= \cdot 0194{,}1447{,}00a, & \sigma_{x_9-x_8} &= \cdot 0190{,}3818{,}39a,\\
\bar{x}_{10} - \bar{x}_9 &= \cdot 0176{,}1683{,}39a, & \sigma_{x_{10}-x_9} &= \cdot 0173{,}0804{,}87a,\\
\bar{x}_{11} - \bar{x}_{10} &= \cdot 0161{,}4876{,}44a, & \sigma_{x_{11}-x_{10}} &= \cdot 0158{,}8976{,}36a.
\end{aligned}$$

The approach to equality between the rank-intervals and their standard deviations is becoming sensible. This equality characterises the exponential curve.

Next, taking the correlations of adjacent intervals, we find

$$\begin{aligned}
r_{x_2-x_1,\,x_3-x_2} &= -\cdot 0779{,}3029, & r_{x_{10}-x_9,\,x_{11}-x_{10}} &= -\cdot 0164{,}5761,\\
r_{x_3-x_2,\,x_4-x_3} &= -\cdot 0506{,}2378, & r_{x_9-x_8,\,x_{10}-x_9} &= -\cdot 0180{,}3372,\\
r_{x_4-x_3,\,x_5-x_4} &= -\cdot 0380{,}8816, & r_{x_8-x_7,\,x_9-x_8} &= -\cdot 0200{,}0923,\\
r_{x_5-x_4,\,x_6-x_5} &= -\cdot 0307{,}8122, & r_{x_7-x_6,\,x_8-x_7} &= -\cdot 0225{,}5633,
\end{aligned}$$

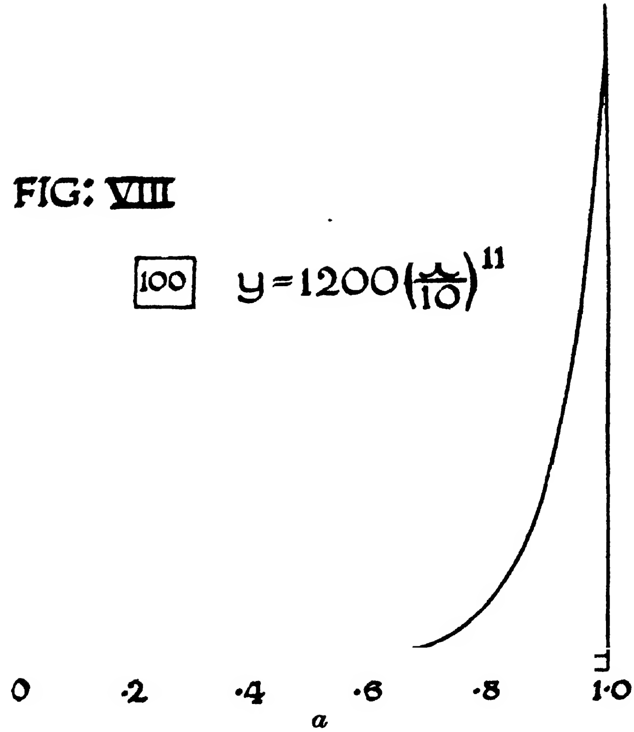
$$r_{x_6-x_5,\,x_7-x_6} = -\cdot 0259{,}6816.$$

The correlations of the interrank intervals are clearly growing smaller and smaller, although the first two intervals have still a considerably higher correlation than the last two.

Illustration VIII. The last curve I propose to take to illustrate *Type IX* is

$$\text{Curve (viii)}: \; y = 1200\left(\frac{x}{a}\right)^{11}. \text{ See Fig. VIII, p. 270.}$$

Here $\sigma = \cdot 0772{,}7776a, \quad \beta_1 = 2\cdot 509{,}630, \quad \beta_2 = 6\cdot 154{,}167.$



The mean rank-variates and their variabilities are given by

$$\begin{aligned}
\bar{x}_1 &= \cdot 7815{,}5832{,}80a, & \sigma_{x_1} &= \cdot 0767{,}3475{,}60a,\\
\bar{x}_2 &= \cdot 8466{,}8818{,}87a, & \sigma_{x_2} &= \cdot 0512{,}6104{,}27a,\\
\bar{x}_3 &= \cdot 8819{,}6686{,}32a, & \sigma_{x_3} &= \cdot 0400{,}2604{,}35a,\\
\bar{x}_4 &= \cdot 9064{,}6594{,}28a, & \sigma_{x_4} &= \cdot 0330{,}2853{,}39a,\\
\bar{x}_5 &= \cdot 9253{,}5064{,}99a, & \sigma_{x_5} &= \cdot 0279{,}2320{,}31a,\\
\bar{x}_6 &= \cdot 9407{,}7316{,}07a, & \sigma_{x_6} &= \cdot 0238{,}2940{,}64a,\\
\bar{x}_7 &= \cdot 9538{,}3945{,}46a, & \sigma_{x_7} &= \cdot 0203{,}1961{,}98a,\\
\bar{x}_8 &= \cdot 9651{,}9468{,}62a, & \sigma_{x_8} &= \cdot 0171{,}3990{,}39a,\\
\bar{x}_9 &= \cdot 9752{,}4879{,}76a, & \sigma_{x_9} &= \cdot 0141{,}0005{,}29a,\\
\bar{x}_{10} &= \cdot 9842{,}7887{,}90a, & \sigma_{x_{10}} &= \cdot 0109{,}9776{,}40a,\\
\bar{x}_{11} &= \cdot 9924{,}8120{,}30a, & \sigma_{x_{11}} &= \cdot 0074{,}6247{,}54a.
\end{aligned}$$

For the interrank intervals we have

$$\begin{aligned}
\bar{x}_2 - \bar{x}_1 &= \cdot 0651{,}2986{,}07a, & \sigma_{x_2-x_1} &= \cdot 0605{,}3750{,}06a,\\
\bar{x}_3 - \bar{x}_2 &= \cdot 0352{,}7867{,}45a, & \sigma_{x_3-x_2} &= \cdot 0339{,}6729{,}41a,\\
\bar{x}_4 - \bar{x}_3 &= \cdot 0244{,}9907{,}95a, & \sigma_{x_4-x_3} &= \cdot 0238{,}7816{,}83a,\\
\bar{x}_5 - \bar{x}_4 &= \cdot 0188{,}8470{,}71a, & \sigma_{x_5-x_4} &= \cdot 0185{,}2035{,}02a,\\
\bar{x}_6 - \bar{x}_5 &= \cdot 0154{,}2251{,}08a, & \sigma_{x_6-x_5} &= \cdot 0151{,}8161{,}46a,
\end{aligned}$$

$$\bar{x}_7 - \bar{x}_6 = \cdot 0130{,}6629{,}39a, \quad \sigma_{x_7 - x_6} = \cdot 0128{,}9444{,}19a,$$
$$\bar{x}_8 - \bar{x}_7 = \cdot 0113{,}5523{,}16a, \quad \sigma_{x_8 - x_7} = \cdot 0112{,}2599{,}78a,$$
$$\bar{x}_9 - \bar{x}_8 = \cdot 0100{,}5411{,}13a, \quad \sigma_{x_9 - x_8} = \cdot 0099{,}5309{,}10a,$$
$$\bar{x}_{10} - \bar{x}_9 = \cdot 0090{,}3008{,}15a, \quad \sigma_{x_{10} - x_9} = 0089{,}4874{,}05a,$$
$$\bar{x}_{11} - \bar{x}_{10} = \cdot 0082{,}0232{,}40a, \quad \sigma_{x_{11} - x_{10}} = \cdot 0081{,}3527{,}77a.$$

Here is even a closer approach to the equality of the interrank interval and its variability. For the correlations of the successive rank-intervals we find

$$r_{x_2 - x_1,\, x_3 - x_2} = - \cdot 0406{,}4928, \quad r_{x_{10} - x_9,\, x_{11} - x_{10}} = - \cdot 0037{,}4562,$$
$$r_{x_3 - x_2,\, x_4 - x_3} = - \cdot 0263{,}4168, \quad r_{x_9 - x_8,\, x_{10} - x_9} = - \cdot 0055{,}0072,$$
$$r_{x_4 - x_3,\, x_5 - x_4} = - \cdot 0193{,}5992, \quad r_{x_8 - x_7,\, x_9 - x_8} = - \cdot 0073{,}7364,$$
$$r_{x_5 - x_4,\, x_6 - x_5} = - \cdot 0150{,}1790, \quad r_{x_7 - x_6,\, x_8 - x_7} = - \cdot 0094{.}5803,$$
$$r_{x_6 - x_5,\, x_7 - x_6} = - \cdot 0119{,}1094.$$

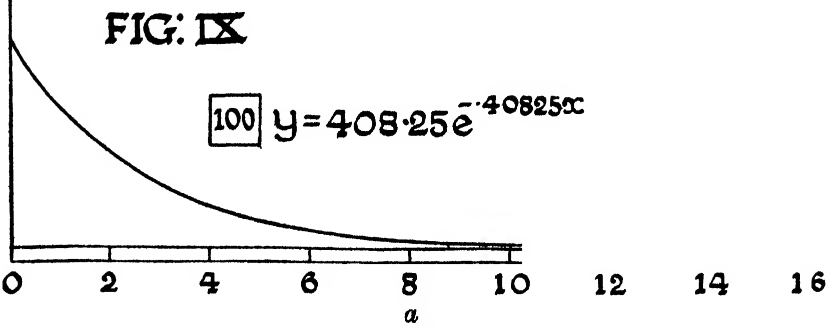
With regard to the interrank interval correlations we see that they are still sensibly negative and very small. The surprising part is that so little correlation should exist between even adjacent rank-intervals, and how close in magnitude the interval and its standard deviation are, right away from the rectangular to the exponential population.

We now reach the Exponential Curve itself. Here the range becomes unlimited in one direction, and with the following curves a is not a physical character of the distribution, it denotes the distance from the origin to the start of the curve, and we shall replace it in our examples by its value in terms of the standard deviation.

Illustration IX. Type X. This is the Exponential Curve:

$$\text{Curve (ix)}: \; y = \frac{N}{\sigma} e^{-x/\sigma}, \; (\beta_1 = 4, \beta_2 = 9),$$

the transition curve from Type IX with finite range to Type XI with infinite range. This type has been fully discussed theoretically in the first section of the present paper. I add a graph here for the purpose of comparison with adjacent curves*.



* For the purpose of plotting σ was taken $= 2 \cdot 449{,}4897$; the graph is therefore directly comparable with the last curve ($y = 600\,(10/x)^7$) of our series.

To carry out the sequence of our curves we give the results for a sample of 11 taken from this exponential curve. We have

$$\bar{x}_1 = \cdot 0909{,}0909{,}09\sigma, \quad \sigma_{x_1} = \cdot 0909{,}0909{,}09\sigma,$$
$$\bar{x}_2 = \cdot 1909{,}0909{,}09\sigma, \quad \sigma_{x_2} = \cdot 1351{,}4607{,}95\sigma,$$
$$\bar{x}_3 = \cdot 3020{,}2020{,}20\sigma, \quad \sigma_{x_3} = \cdot 1749{,}5754{,}29\sigma,$$
$$\bar{x}_4 = \cdot 4270{,}2020{,}20\sigma, \quad \sigma_{x_4} = \cdot 2150{,}2358{,}43\sigma,$$
$$\bar{x}_5 = \cdot 5698{,}7734{,}49\sigma, \quad \sigma_{x_5} = \cdot 2581{,}5364{,}63\sigma,$$
$$\bar{x}_6 = \cdot 7365{,}4401{,}15\sigma, \quad \sigma_{x_6} = \cdot 3072{,}8013{,}74\sigma,$$
$$\bar{x}_7 = \cdot 9365{,}4401{,}15\sigma, \quad \sigma_{x_7} = \cdot 3666{,}3480{,}86\sigma,$$
$$\bar{x}_8 = 1\cdot 1865{,}4401{,}15\sigma, \quad \sigma_{x_8} = \cdot 4437{,}5791{,}02\sigma,$$
$$\bar{x}_9 = 1\cdot 5198{,}7734{,}49\sigma, \quad \sigma_{x_9} = \cdot 5550{,}0648{,}10\sigma,$$
$$\bar{x}_{10} = 2\cdot 0198{,}7734{,}49\sigma, \quad \sigma_{x_{10}} = \cdot 7470{,}1552{,}46\sigma,$$
$$\bar{x}_{11} = 3\cdot 0198{,}7734{,}49\sigma, \quad \sigma_{x_{11}} = 1\cdot 2482{,}1159{,}82\sigma.$$

Now these results cannot be compared directly with those on p. 270 for the curve $y = y_0\left(\frac{x}{a}\right)^{11}$. We have first to notice that the rank-variates are measured in the present curve from the maximum frequency end of the range, and are in terms of the range a and not the standard deviation σ. Accordingly for comparison we must subtract each rank-variate from the range a, renumber them in reverse order, and substitute $14\cdot 0416{,}0480\sigma$ for a. We then find for $y = y_0(x/a)^{11}$ the two series

$$\bar{x}_1' = \cdot 1055{,}7597{,}60\sigma, \quad \sigma_{x_1'} = \cdot 1047{,}8513{,}04\sigma,$$
$$\bar{x}_2' = \cdot 2207{,}4976{,}81\sigma, \quad \sigma_{x_2'} = \cdot 1544{,}2625{,}58\sigma,$$
$$\bar{x}_3' = \cdot 3475{,}4660{,}24\sigma, \quad \sigma_{x_3'} = \cdot 1979{,}8737{,}05\sigma,$$
$$\bar{x}_4' = \cdot 4887{,}2246{,}13\sigma, \quad \sigma_{x_4'} = \cdot 2406{,}7175{,}69\sigma,$$
$$\bar{x}_5' = \cdot 6481{,}6813{,}59\sigma, \quad \sigma_{x_5'} = \cdot 2853{,}2007{,}09\sigma,$$
$$\bar{x}_6' = \cdot 8316{,}3987{,}10\sigma, \quad \sigma_{x_6'} = \cdot 3346{,}0310{,}73\sigma,$$
$$\bar{x}_7' = 1\cdot 0481{,}9667{,}27\sigma, \quad \sigma_{x_7'} = \cdot 3920{,}8658{,}27\sigma,$$
$$\bar{x}_8' = 1\cdot 313\dot{3}{,}6826{,}65\sigma, \quad \sigma_{x_8'} = \cdot 4637{,}7362{,}0\dot{1}\sigma,$$
$$\bar{x}_9' = 1\cdot 6573{,}4660{,}25\sigma, \quad \sigma_{x_9'} = \cdot 5620{,}2988{,}45\sigma,$$
$$\bar{x}_{10}' = 2\cdot 1527{,}4386{,}40\sigma, \quad \sigma_{x_{10}'} = \cdot 7197{,}8730{,}32\sigma,$$
$$\bar{x}_{11}' = 3\cdot 0672{,}7163{,}01\sigma, \quad \sigma_{x_{11}'} = 1\cdot 0774{,}7911{,}82\sigma.$$

These values, although not very close to those above—for the curve with 11 as power is only a very rough approach to the infinity of the exponential—yet suffice to indicate that the rank-variates and their standard deviations are approaching the exponential values.

Turning now to the interrank intervals we find

$$\bar{x}_2 - \bar{x}_1 = \cdot 1000{,}0000{,}00\sigma, \quad = \sigma_{x_2-x_1},$$
$$\bar{x}_3 - \bar{x}_2 = \cdot 1111{,}1111{,}11\sigma, \quad = \sigma_{x_3-x_2},$$
$$\bar{x}_4 - \bar{x}_3 = \cdot 1250{,}0000{,}00\sigma, \quad = \sigma_{x_4-x_3},$$
$$\bar{x}_5 - \bar{x}_4 = \cdot 1428{,}5714{,}29\sigma, \quad = \sigma_{x_5-x_4},$$
$$\bar{x}_6 - \bar{x}_5 = \cdot 1666{,}6666{,}67\sigma, \quad = \sigma_{x_6-x_5},$$
$$\bar{x}_7 - \bar{x}_6 = \cdot 2000{,}0000{,}00\sigma, \quad = \sigma_{x_7-x_6},$$
$$\bar{x}_8 - \bar{x}_7 = \cdot 2500{,}0000{,}00\sigma, \quad = \sigma_{x_8-x_7},$$
$$\bar{x}_9 - \bar{x}_8 = \cdot 3333{,}3333{,}33\sigma, \quad = \sigma_{x_9-x_8},$$
$$\bar{x}_{10} - \bar{x}_9 = \cdot 5000{,}0000{,}00\sigma,$$
$$\bar{x}_{11} - \bar{x}_{10} = 1\cdot 0000{,}0000{,}00\sigma, \quad = \sigma_{x_{11}-x_{10}}.$$

The corresponding values for $y = y_0 (x/a)^{11}$ are

$$\bar{x}_2' - \bar{x}_1' = \cdot 1151{,}7379{,}20\sigma, \quad \sigma_{x_2'-x_1'} = \cdot 1142{,}3235{,}44\sigma,$$
$$\bar{x}_3' - \bar{x}_2' = \cdot 1267{,}9683{,}57\sigma, \quad \sigma_{x_3'-x_2'} = \cdot 1256{,}5467{,}76\sigma,$$
$$\bar{x}_4' - \bar{x}_3' = \cdot 1411{,}7585{,}75\sigma, \quad \sigma_{x_4'-x_3'} = \cdot 1397{,}5737{,}04\sigma,$$
$$\bar{x}_5' - \bar{x}_4' = \cdot 1594{,}4567{,}45\sigma, \quad \sigma_{x_5'-x_4'} = \cdot 1576{,}3102{,}46\sigma,$$
$$\bar{x}_6' - \bar{x}_5' = \cdot 1834{,}7173{,}51\sigma, \quad \sigma_{x_6'-x_5'} = \cdot 1810{,}5865{,}73\sigma,$$
$$\bar{x}_7' - \bar{x}_6' = \cdot 2165{,}5680{,}17\sigma, \quad \sigma_{x_7'-x_6'} = \cdot 2131{,}7423{,}24\sigma,$$
$$\bar{x}_8' - \bar{x}_7' = \cdot 2651{,}7159{,}39\sigma, \quad \sigma_{x_8'-x_7'} = \cdot 2600{,}5543{,}83\sigma,$$
$$\bar{x}_9' - \bar{x}_8' = \cdot 3440{,}0639{,}23\sigma, \quad \sigma_{x_9'-x_8'} = \cdot 3352{,}8780{,}26\sigma,$$
$$\bar{x}_{10}' - \bar{x}_9' = \cdot 4953{,}6920{,}52\sigma, \quad \sigma_{x_{10}'-x_9'} = \cdot 4769{,}5531{,}99\sigma,$$
$$\bar{x}_{11}' - \bar{x}_{10}' = \cdot 9145{,}2776{,}46\sigma, \quad \sigma_{x_{11}'-x_{10}'} = \cdot 8500{,}4365{,}90\sigma.$$

These series show the approach to the exponential values and at the same time indicate the influence of the limited range on the higher intervals at the tail. While for the exponential curve the interval correlations are all zero, we see them approaching zero for the 11th power curve (p. 271). To compare with curves beyond the exponential the correlations on p. 275 must be read in reverse order, $r_{x_{10}-x_9,\, x_{11}-x_{10}}$ being read as $r_{x_2'-x_1',\, x_3'-x_2'}$.

Illustration X. Type XI.

$$\text{Curve (x)}: \; y = y_0 \left(\frac{x}{a}\right)^{-20}, \text{ with range from } x = a \text{ to } \infty.$$

Here $\quad \sigma = \cdot 0587{,}3268{,}22a$*, $\beta_1 = 5\cdot 592{,}1053$, $\beta_2 = 12\cdot 347{,}3684$,

and $\quad y_0 = 19N/a.$

The curve is shown on the same scale as the previous curves in Fig. X.

* $a = 17\cdot 026{,}2954\sigma$ is the value to be used in passing from the Γ-function expressions with a to those in σ.

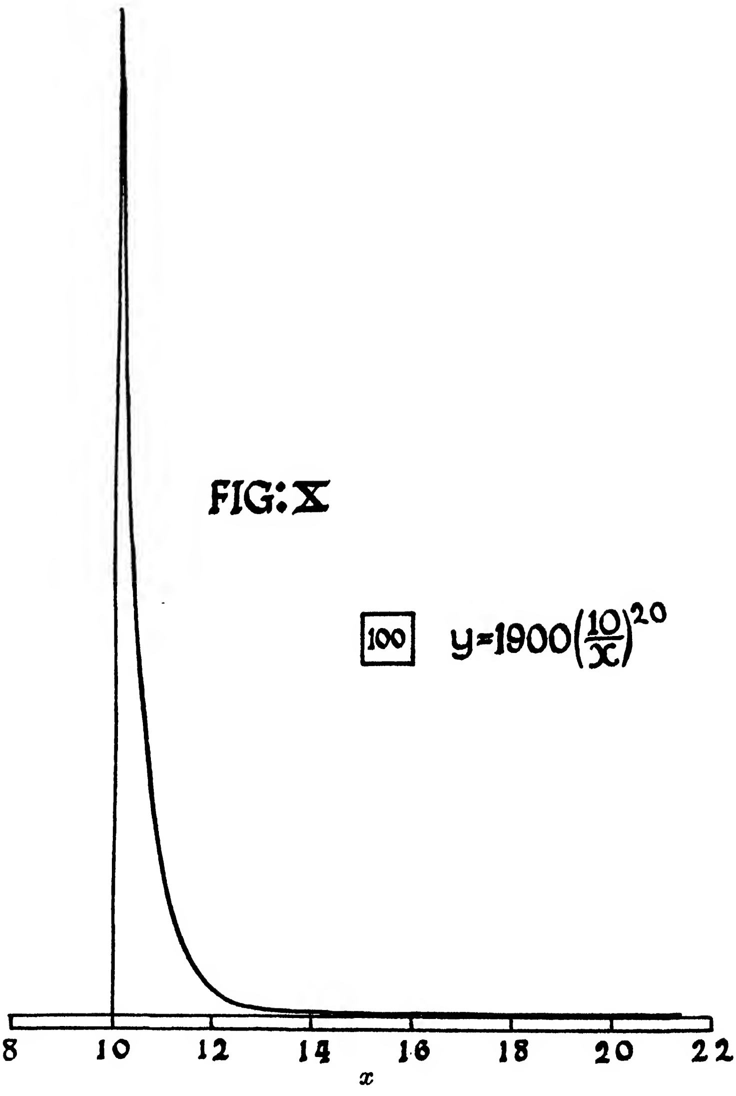


FIG: X

We give first the rank-variates measured no longer from the origin of x, but reduced to the mode, and given in terms of σ, not a.

$$\bar{x}_1' = \cdot0818{,}5718{,}94\sigma, \quad \sigma_{x_1'} = \cdot0822{,}5171{,}24\sigma,$$
$$\bar{x}_2' = \cdot1723{,}7651{,}49\sigma, \quad \sigma_{x_2'} = \cdot1229{,}5625{,}72\sigma,$$
$$\bar{x}_3' = \cdot2735{,}4517{,}36\sigma, \quad \sigma_{x_3'} = \cdot1601{,}6667{,}75\sigma,$$
$$\bar{x}_4' = \cdot3881{,}1365{,}45\sigma, \quad \sigma_{x_4'} = \cdot1982{,}3306{,}93\sigma,$$
$$\bar{x}_5' = \cdot5200{,}4099{,}52\sigma, \quad \sigma_{x_5'} = \cdot2399{,}3092{,}14\sigma,$$
$$\bar{x}_6' = \cdot6753{,}1830{,}81\sigma, \quad \sigma_{x_6'} = \cdot2883{,}3392{,}43\sigma,$$
$$\bar{x}_7' = \cdot8636{,}3334{,}75\sigma, \quad \sigma_{x_7'} = \cdot3480{,}6540{,}02\sigma,$$

$$\bar{x}_8' = 1\cdot1021,6573,24\sigma, \quad \sigma_{x_8'} = \cdot4276,2048,79\sigma,$$
$$\bar{x}_9' = 1\cdot4258,8825,15\sigma, \quad \sigma_{x_9'} = \cdot5470,8444,81\sigma,$$
$$\bar{x}_{10}' = 1\cdot9245,9591,83\sigma, \quad \sigma_{x_{10}'} = \cdot7597,3959,28\sigma,$$
$$\bar{x}_{11}' = 2\cdot9774,2321,29\sigma, \quad \sigma_{x_{11}'} = 1\cdot3725,7470,15\sigma.$$

The tendency when compared with the exponential curve is to draw the ranked individuals slightly towards mediocrity, and to reduce the variability of the lowest ranks while increasing that of the two highest ranks. Both these tendencies will be found still further emphasised in the following curve, $y = y_0(a/x)^7$.

We turn next to the rank-intervals and find*

$$\bar{x}_2 - \bar{x}_1 = \cdot0905,1932,59\sigma, \quad \sigma_{x_2-x_1} = \cdot0910,0163,12\sigma,$$
$$\bar{x}_3 - \bar{x}_2 = \cdot1011,6865,83\sigma, \quad \sigma_{x_3-x_2} = \cdot1017,7070,00\sigma,$$
$$\bar{x}_4 - \bar{x}_3 = \cdot1145,6848,06\sigma, \quad \sigma_{x_4-x_3} = \cdot1153,3956,09\sigma,$$
$$\bar{x}_5 - \bar{x}_4 = \cdot1319,2734,13\sigma, \quad \sigma_{x_5-x_4} = \cdot1329,4770,10\sigma,$$
$$\bar{x}_6 - \bar{x}_5 = \cdot1552,7731,32\sigma, \quad \sigma_{x_6-x_5} = \cdot1566,8661,51\sigma,$$
$$\bar{x}_7 - \bar{x}_6 = \cdot1883,1503,95\sigma, \quad \sigma_{x_7-x_6} = \cdot1903,7912,02\sigma,$$
$$\bar{x}_8 - \bar{x}_7 = \cdot2385,3238,33\sigma, \quad \sigma_{x_8-x_7} = \cdot2418,2458,76\sigma,$$
$$\bar{x}_9 - \bar{x}_8 = \cdot3237,2252,02\sigma, \quad \sigma_{x_9-x_8} = \cdot3297,3592,92\sigma,$$
$$\bar{x}_{10} - \bar{x}_9 = \cdot4987,0766,63\sigma, \quad \sigma_{x_{10}-x_9} = \cdot5128,1004,50\sigma,$$
$$\bar{x}_{11} - \bar{x}_{10} = 1\cdot0528,2729,54\sigma, \quad \sigma_{x_{11}-x_{10}} = 1\cdot1147,3011,17\sigma.$$

These show the great increase of interval and of its variability in the last interval at the tail. The Galton Ratio is over three, and we obtain much the same value of division of prize money, 75 °/。 and 25 °/。, as determined for the normal curve.

The adjacent interval correlations are as follows:

$$r_{x_2-x_1,\, x_3-x_2} = \cdot0052,8264, \quad r_{x_{10}-x_9,\, x_{11}-x_{10}} = \cdot0263,4030,$$
$$r_{x_3-x_2,\, x_4-x_3} = \cdot0058,9360, \quad r_{x_9-x_8,\, x_{10}-x_9} = \cdot0179,0025,$$
$$r_{x_4-x_3,\, x_5-x_4} = \cdot0066,5633, \quad r_{x_8-x_7,\, x_9-x_8} = \cdot0134,5797,$$
$$r_{x_5-x_4,\, x_6-x_5} = \cdot0077,7444, \quad r_{x_7-x_6,\, x_8-x_7} = \cdot0107,5295,$$
$$r_{x_6-x_5,\, x_7-x_6} = \cdot0089,5670.$$

We have here a very low series of adjacent interval correlations, although we have departed, to judge by this curve's β_1, and specially its β_2, to a considerable distance from the exponential curve.

Illustration XI. Type XI.

$$\text{Curve (xi)}: \; y = y_0\left(\frac{x}{a}\right)^{-7}, \text{ with range from } x = a \text{ to } x = \infty.$$

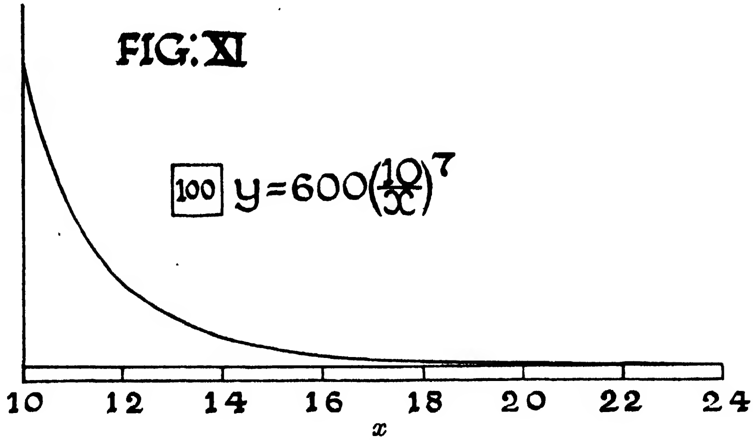
Here $\sigma = a \times \cdot2449,4897$†, $\beta_1 = 14\cdot518,5185$, $\beta_2 = 38\cdot666,6667$

and $$y_0 = 6N/a.$$

* In the case of the intervals we need not distinguish between $x_{q+1} - x_q$ and $x'_{q+1} - x'_q$.

† Conversely $a = 4\cdot0824,8290\sigma$.

This curve is graphed in Fig. XI.



The following values were obtained for a sample of 11 from this curve as parent population:

$$\begin{aligned}
\bar{x}_1 &= 1{\cdot}0153{,}8461{,}54a, & \sigma_{x_1} &= {\cdot}0156{,}2315{,}08a,\\
\bar{x}_2 &= 1{\cdot}0325{,}9452{,}41a, & \sigma_{x_2} &= {\cdot}0236{,}4093{,}03a,\\
\bar{x}_3 &= 1{\cdot}0520{,}7743{,}97a, & \sigma_{x_3} &= {\cdot}0312{,}1814{,}98a,\\
\bar{x}_4 &= 1{\cdot}0744{,}6206{,}60a, & \sigma_{x_4} &= {\cdot}0392{,}4032{,}14a,\\
\bar{x}_5 &= 1{\cdot}1006{,}6845{,}79a, & \sigma_{x_5} &= {\cdot}0483{,}5190{,}47a,\\
\bar{x}_6 &= 1{\cdot}1321{,}1612{,}81a, & \sigma_{x_6} &= {\cdot}0593{,}5116{,}22a,\\
\bar{x}_7 &= 1{\cdot}1711{,}5461{,}53a, & \sigma_{x_7} &= {\cdot}0735{,}3252{,}88a,\\
\bar{x}_8 &= 1{\cdot}2220{,}7438{,}12a, & \sigma_{x_8} &= {\cdot}0934{,}1902{,}87a,\\
\bar{x}_9 &= 1{\cdot}2939{,}6110{,}95a, & \sigma_{x_9} &= {\cdot}1250{,}2667{,}20a,\\
\bar{x}_{10} &= 1{\cdot}4115{,}9393{,}76a, & \sigma_{x_{10}} &= {\cdot}1880{,}5047{,}50a,\\
\bar{x}_{11} &= 1{\cdot}6939{,}1272{,}52a, & \sigma_{x_{11}} &= {\cdot}4154{,}5194{,}48a.
\end{aligned}$$

Replacing a by σ and measuring from the start of the curve, not from the origin, we find the following series:

$$\begin{aligned}
\bar{x}_1' &= {\cdot}0628{,}0742{,}94\sigma, & \sigma_{x_1'} &= {\cdot}0637{,}8124{,}60\sigma,\\
\bar{x}_2' &= {\cdot}1330{,}6658{,}73\sigma, & \sigma_{x_2'} &= {\cdot}0965{,}1369{,}37\sigma,\\
\bar{x}_3' &= {\cdot}2126{,}0525{,}71\sigma, & \sigma_{x_3'} &= {\cdot}1274{,}4756{,}27\sigma,\\
\bar{x}_4' &= {\cdot}3039{,}9011{,}11\sigma, & \sigma_{x_4'} &= {\cdot}1601{,}9794{,}11\sigma,\\
\bar{x}_5' &= {\cdot}4109{,}7725{,}79\sigma, & \sigma_{x_5'} &= {\cdot}1973{,}9582{,}41\sigma,\\
\bar{x}_6' &= {\cdot}5393{,}6183{,}38\sigma, & \sigma_{x_6'} &= {\cdot}2423{,}0010{,}48\sigma,\\
\bar{x}_7' &= {\cdot}6987{,}3579{,}02\sigma, & \sigma_{x_7'} &= {\cdot}3001{,}9529{,}14\sigma,\\
\bar{x}_8' &= {\cdot}9066{,}1486{,}38\sigma, & \sigma_{x_8'} &= {\cdot}3813{,}8158{,}72\sigma,\\
\bar{x}_9' &= 1{\cdot}2000{,}9120{,}28\sigma, & \sigma_{x_9'} &= {\cdot}5104{,}1925{,}05\sigma,\\
\bar{x}_{10}' &= 1{\cdot}6803{,}2521{,}20\sigma, & \sigma_{x_{10}'} &= {\cdot}7677{,}1284{,}85\sigma,\\
\bar{x}_{11}' &= 2{\cdot}8328{,}8683{,}47\sigma, & \sigma_{x_{11}'} &= 1{\cdot}6960{,}7546{,}04\sigma.
\end{aligned}$$

It will be seen at once how marked is becoming the influence of the tail in separating both variates and variabilities of the highest ranks.

We have then the following values for the rank-intervals and their standard deviations*:

$$
\begin{array}{llll}
\bar{x}_2 - \bar{x}_1 = & \cdot 0702{,}5915{,}80\sigma, & \sigma_{x_2-x_1} = & \cdot 0714{,}7689{,}04\sigma, \\
\bar{x}_3 - \bar{x}_2 = & \cdot 0795{,}3866{,}94\sigma, & \sigma_{x_3-x_2} = & \cdot 0810{,}9551{,}65\sigma, \\
\bar{x}_4 - \bar{x}_3 = & \cdot 0913{,}8485{,}45\sigma, & \sigma_{x_4-x_3} = & \cdot 0934{,}3079{,}12\sigma, \\
\bar{x}_5 - \bar{x}_4 = & \cdot 1069{,}8714{,}68\sigma, & \sigma_{x_5-x_4} = & \cdot 1097{,}7184{,}92\sigma, \\
\bar{x}_6 - \bar{x}_5 = & \cdot 1283{,}8457{,}58\sigma, & \sigma_{x_6-x_5} = & \cdot 1323{,}5426{,}59\sigma, \\
\bar{x}_7 - \bar{x}_6 = & \cdot 1593{,}7395{,}64\sigma, & \sigma_{x_7-x_6} = & \cdot 1654{,}0541{,}75\sigma, \\
\bar{x}_8 - \bar{x}_7 = & \cdot 2078{,}7907{,}36\sigma, & \sigma_{x_8-x_7} = & \cdot 2179{,}4132{,}88\sigma, \\
\bar{x}_9 - \bar{x}_8 = & \cdot 2934{,}7633{,}90\sigma, & \sigma_{x_9-x_8} = & \cdot 3129{,}9186{,}31\sigma, \\
\bar{x}_{10} - \bar{x}_9 = & \cdot 4802{,}3400{,}92\sigma, & \sigma_{x_{10}-x_9} = & \cdot 5305{,}5302{,}65\sigma, \\
\bar{x}_{11} - \bar{x}_{10} = & 1\cdot 1525{,}6162{,}23\sigma, & \sigma_{x_{11}-x_{10}} = & 1\cdot 4323{,}1831{,}01\sigma.
\end{array}
$$

Here the Galton Ratio is almost 3·4, and the high values of the highest interval and its variability are conspicuous.

Finally, the correlations, all plus, are as follows:

$$
\begin{array}{ll}
r_{x_2-x_1,\, x_3-x_2} = \cdot 0168{,}5446, & r_{x_{10}-x_9,\, x_{11}-x_{10}} = \cdot 0803{,}1649, \\
r_{x_3-x_2,\, x_4-x_3} = \cdot 0189{,}6107, & r_{x_9-x_8,\, x_{10}-x_9} = \cdot 0583{,}1446, \\
r_{x_4-x_3,\, x_5-x_4} = \cdot 0215{,}8128, & r_{x_8-x_7,\, x_9-x_8} = \cdot 0443{,}3855. \\
r_{x_5-x_4,\, x_6-x_5} = \cdot 0249{,}2747, & r_{x_7-x_6,\, x_8-x_7} = \cdot 0354{,}3929, \\
\multicolumn{2}{c}{r_{x_6-x_5,\, x_7-x_6} = \cdot 0293{,}4601.}
\end{array}
$$

These are all small and confirm our previous conclusion that we can move a long way from the (β_1, β_2) of the exponential curve and yet find that the correlations between adjacent rank-intervals, and therefore *a fortiori* between non-adjacent rank-intervals have not reached any very sensible values.

Summary.

It may be asked: Why in this paper have so much labour and space been devoted to the numerical illustration of a sample with definite size from a particular category of frequency curves? The answer is threefold.

(i) This system of curves has a particularly practical value. It is not only that the Exponential Curve as one of the series is very important as the curve describing the random distribution of the occurrence of events in time or space. That curve is deduced by the conception of a happening at a point of time or a point of space. But very often an occurrence cannot so happen. It happens within a certain unit of time, or a certain unit of space. It is a purely mathematical conception to find a "thing" at a point in space, or to mark a "happening" occurring at an instant of time! You may try to make a record of this kind and you will find it impossible;

* In the case of the intervals we need not distinguish between $x_{q+1} - x_q$ and $x'_{q+1} - x'_q$.

you can only measure your space or time in definite units, however small, and more generally the happening did or did not occur within a quite appreciable measure of space or time. In such cases we cannot proceed as Whitworth has done by taking our interval indefinitely small, and thus reaching the exponential limit. The actual frequency curve then falls into the biquadratic series of curves, but is not the exponential curve.

(ii) It is the writer's opinion, and this is based on a fairly long statistical experience, that mathematical formulae convey little to the mind, until they have been numerically illustrated, and this illustration brings out new points, which the formulae actually contain, but which might have been passed over, but for their numerical illustration. Certain broad principles of this kind arise from our work. For example:

(*a*) The correlation between adjacent rank-variates is high, but the correlation between adjacent (and *a fortiori* non-adjacent) rank-intervals is small, and for many purposes negligible.

(*b*) The partial correlation of any two rank-variates, or any two rank-intervals for a constant rank-variate or a constant rank-interval lying between them is zero.

(*c*) The order of the variabilities of rank-intervals as measured by their standard deviations is much the same as the order of the intervals themselves. There is equality in the case of the exponential curve, and this property extends approximately for a considerable range on either side of it.

(*d*) Galton's Ratio, namely 2 to 1, for the ratio of the first rank-interval to the second in the case of the end of the curve with lesser frequency is *approximately* true for a large number of curves.

(*e*) In samples from a curve of finite range the correlations of adjacent interrank intervals are negative; in samples of one with unlimited range they are positive.

(*f*) In cases where there is much predominance of mediocrity, the interval between the first and second ranks may be ten or more times the interval between mediocre individuals. This is but a special illustration of the great principle (which ought to be generally recognised, but frequently is not) that differences in physical or mental ability between specially able individuals will invariably be found to be much greater than those between ordinary individuals. Several characteristics of the so-called "genius" are involved in this principle.

(iii) The statistical characters of the curves dealt with have a very wide range. Thus

Curve (i): $\beta_1 = 2\cdot7500$, $\beta_2 = 4\cdot7143$,
Curve (ii): $\beta_1 = 1\cdot4286$, $\beta_2 = 2\cdot1428$,
Curve (iii): $\beta_1 = 0\cdot0000$, $\beta_2 = 1\cdot8000$,
Curve (iv): $\beta_1 = 0\cdot0067$, $\beta_2 = 1\cdot8316$,
Curve (v): $\beta_1 = 0\cdot0494$, $\beta_2 = 1\cdot9281$,

Curve (vi): $\beta_1 = 0{\cdot}3200$, $\beta_2 = 2{\cdot}4000$,
Curve (vii): $\beta_1 = 1{\cdot}6461$, $\beta_2 = 4{\cdot}6222$,
Curve (viii): $\beta_1 = 2{\cdot}5096$, $\beta_2 = 6{\cdot}1542$,
Curve (ix): $\beta_1 = 4{\cdot}0000$, $\beta_2 = 9{\cdot}0000$,
Curve (x): $\beta_1 = 5{\cdot}5921$, $\beta_2 = 12{\cdot}3474$,
Curve (xi): $\beta_1 = 14{\cdot}5185$, $\beta_2 = 38{\cdot}6667$.

Many of these curves are of a most skew character, and they are very widely spread over the β_1, β_2 plane. It is not suggested that the ranking characters of *all* frequency curves can be judged by these illustrations, but it is certain that the general properties we have stated for these curves are likely to hold for curves with their β_1, β_2 not very far from the biquadratic, and that some of these properties may possibly hold for samples from all continuous frequency curves.

Lastly, I venture to think that a more intensive study, analytical and experimental, of the individuals in a sample, rather than of the statistical constants of the sample as a whole may suggest results of considerable practical value.

MISCELLANEA.

(1) A Comparison of the Accuracy of Two Types of Quadrature Formulae.

By E. S. MARTIN.

The object of this paper is to examine and compare the accuracy of two types of approximate quadrature formulae as applied to commonly occurring frequency curves*.

Let the area between the curve whose equation is $y=f(x)$, and the axis of x be divided into strips by a series of ordinates distant h apart, and the mid-ordinates of these strips calculated; then it is required to calculate the area of each strip.

Let x_r be the abscissa corresponding to the mid-ordinate y_r of any strip; by fitting a "parabola" of the form

$$y=a+bx+cx^2+dx^3+ex^4$$

to the five consecutive mid-ordinates whose middle one is y_r, the following formula is obtained for the area of the strip with mid-ordinate y_r:

$$n_r=\int_{x_r-h/2}^{x_r+h/2} y\,dx=\frac{h}{5760}\left[5178y_r+308\,(y_{r-1}+y_{r+1})-17\,(y_{r-2}+y_{r+2})\right] \text{ ....(i).}$$

If the curve has a finite terminal, in order to calculate the areas of the two terminal strips, we may either (a) assign zero values to one or two of the ordinates in the above formula, or (b) fit a "parabola" to the five terminal mid-ordinates. If (b) is adopted, the following formulae are obtained for the first two strips, with equivalent formulae for the last two strips:

$$\left.\begin{aligned} n_1&=\frac{h}{5760}\left[6463y_1-2092y_2+2298y_3-1132y_4+223y_5\right]\\ n_2&=\frac{h}{5760}\left[223y_1+5348y_2+138y_3+68y_4-17y_5\right]\end{aligned}\right\} \text{ ....(i)}'.$$

The second type of formula is obtained on the assumption that $f(x)$ can be expanded by Taylor's Theorem, and has the general form

$$n_r=h\left[f(x_r)+\frac{h^2}{24}f''(x_r)\right] \text{ ....(ii),}$$

neglecting terms in h^5, etc.

This, of course, requires to be put in a form suitable for calculation for each curve considered. The equations of the curves here tabulated, with the corresponding forms of (ii), are:

$$\text{Normal Curve}\quad y=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\quad (\text{Tables I}^{a-c}),$$

$$n_r=hy\left[1+\frac{h^2}{24}(x^2-1)\right].$$

$$\text{Curve of type}\quad y=\frac{1}{B(m_1+1,\,m_2+1)}x^{m_1}(1-x)^{m_2}\quad (\text{Tables II}^{a-c}),$$

$$n_r=hy\left[1+\frac{h^2}{24}\,\frac{(\overline{m_1+m_2}\,x-m_1)^2(m_1+m_2-1)-m_1m_2}{(m_1+m_2)\,x^2(1-x)^2}\right],$$

* *Biometrika*, Vol. III. pp. 310—312, and *Tables for Statisticians and Biometricians*, Part II. p. xvi.

which for calculation purposes is put in the form

$$n_r = hy\left[1 + h^2\frac{(x-a)(x-\beta)}{\gamma \cdot x^2(1-x)^2}\right],$$

where a, β and γ depend on the constants m_1 and m_2 of the curve.

$$\text{Curve of type} \qquad y = \frac{1}{\Gamma(p+1)}x^p e^{-x} \qquad (\text{Tables III}^{a-c}, \text{IV—VII}),$$

$$n_r = hy\left[1 + \frac{h^2}{24}\frac{(x-p)^2 - p}{x^2}\right].$$

The last two of these curves are respectively Pearson's Type I and Type III frequency curves, in simplest form, with total frequency unity and origin at one finite terminal.

In the tables appended there are set out the mid-ordinate y of each strip, and its corresponding abscissa x; the true area n of each strip; the value $n_{(i)}$ obtained from formula (i); the value $n_{(ii)}$ obtained from formula (ii); and the errors in the last two results. Where the curve has a finite terminal and the two terminal strips are appreciable, there are set out for comparison n, $n_{(i)}$ obtained by putting one or two ordinates zero in (i), $n_{(i)'}$ obtained from (i)′, and $n_{(ii)}$.

In the tables of the Normal Curve the true values n are taken from Sheppard's tables of this curve; in those of the Type I curve they are taken from unpublished tables of the Incomplete B-Function in the Department of Applied Statistics, University College, London; while for the Type III curves the true values are obtained from the published tables of the Incomplete Γ-Function.

An examination of the tables shows that, disregarding terminal strips, there is very little to choose, as regards accuracy, between the formulae (i) and (ii). For an interval h equal to one-half the standard deviation σ of the curve, both formulae give, in about 90 °/$_\circ$ of results, three significant figures correctly, with only small errors in the third figure in the remaining results; in about 70 °/$_\circ$ of results, four significant figures are correct. For $h = \frac{1}{3}\sigma$ or smaller, both formulae can be reasonably relied on to give everywhere four or more significant figures correctly.

With regard to the terminal strips, it is seen that formulae (i)′, specially devised for these strips, give by far the worst results (except in Table III^a); ludicrously bad results are obtained for the actual end strip. Formula (ii) is obviously the more accurate for the terminal strips, in the curves examined. The areas under consideration, however, are small compared with the total area; thus for $h = \frac{1}{2}\sigma$, the most unfavourable case considered is in Table III^a, where the two terminal strips amount together to about 14 °/$_\circ$ of the total area. The total error involved in using formula (i) for these two areas is less than 0·2 °/$_\circ$ of the total area, i.e., less than 1 in a total frequency of 500. For smaller values of h the error is of course much smaller.

Thus it would appear from these tables that formula (i), being as accurate as, and much easier of application than, formula (ii), is the better formula from which to calculate sub-frequencies in the types of distributions considered; it is also accurate enough for most purposes in calculating terminal frequencies.

It is hoped that the tables will indicate the degree of accuracy to be expected for various values of h in the more usual frequency distributions.

TABLE I^a.

Normal Curve: $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.

$h = \frac{1}{2}\sigma = \frac{1}{2}$.

					Error × 10^6	
x	y	n	$n_{(i)}$	$n_{(ii)}$	in $n_{(i)}$	in $n_{(ii)}$
0	·398 942	·197 413	·197 428	·197 393	+15	−20
± ·5	·352 065	·174 666	·174 671	·174 657	+ 5	− 9
±1·0	·241 971	·120 978	·120 968	·120 985	−10	+ 7
±1·5	·129 518	·065 591	·065 583	·065 602	− 8	+11
±2·0	·053 991	·027 835	·027 836	·027 839	+ 1	+ 4
±2·5	·017 528	·009 245	·009 248	·009 243	+ 3	− 2
±3·0	·004 432	·002 403	·002 404	·002 401	+ 1	− 2
±3·5	·000 873	·000 489	·000 488	·000 487	− 1	− 2
±4·0	·000 134	·000 078	·000 077	·000 077	− 1	− 1
±4·5	·000 016	·000 010	·000 010	·000 010	0	0
±5·0	·000 001	·000 001	·000 001	·000 001	0	0

TABLE I^b.

$h = \frac{1}{3}\sigma = \frac{1}{3}$.

					Error × 10^6	
x	y	n	$n_{(i)}$	$n_{(ii)}$	in $n_{(i)}$	in $n_{(ii)}$
0	·398 942	·132 368	·132 369	·132 365	+1	−3
± $\cdot\dot{3}$	·377 384	·125 279	·125 279	·125 277	0	−2
± $\cdot\dot{6}$	·319 448	·106 209	·106 209	·106 209	0	0
±1·0	·241 971	·080 656	·080 655	·080 657	−1	+1
±1·$\dot{3}$	·164 010	·054 865	·054 865	·054 867	0	+2
±1·$\dot{6}$	·099 477	·033 431	·033 430	·033 432	−1	+1
±2·0	·053 991	·018 246	·018 246	·018 247	0	+1
±2·$\dot{3}$	·026 222	·008 920	·008 921	·008 920	+1	0
±2·$\dot{6}$	·011 396	·003 906	·003 907	·003 906	+1	0
±3·0	·004 432	·001 532	·001 532	·001 532	0	0
±3·$\dot{3}$	·001 542	·000 538	·000 538	·000 538	0	0
±3·$\dot{6}$	·000 480	·000 169	·000 169	·000 169	0	0
±4·0	·000 134	·000 048	·000 048	·000 048	0	0
±4·$\dot{3}$	·000 033	·000 012	·000 012	·000 012	0	0
±4·$\dot{6}$	·000 007(5)	·000 003	·000 003	·000 003	0	0
±5·0	·000 001	·000 001	·000 001	·000 001	0	0

TABLE I^c.

$h = \frac{1}{6}\sigma = \frac{1}{6}$.

This table has been calculated and it is found that $n_{(i)}$ and $n_{(ii)}$ are everywhere correct to six decimal places.

TABLE IIᵃ.

Equation of Curve: $y = \dfrac{1}{\cdot 000\,000\,069\,694\,06} x^{48}(1-x)^{4}$.

$h = \cdot 04 = \sigma$ approx.

x	y	n	$n_{(i)}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(ii)}$
·58	·000 002	·000 000	·000 000	·000 000	0	0
·62	·000 032	·000 002	·000 002	·000 002	0	0
·66	·000 418	·000 021	·000 020	·000 021	− 1	0
·70	·004 266	·000 205	·000 202	·000 204	− 3	− 1
·74	·034 659	·001 606	·001 601	·001 599	− 5	− 7
·78	·222 3[illegible]	·009 933	·009 952	·009 913	+ 19	− 20
·82	1·098 908	·047 255	·047 370	·047 250	+ 115	− 5
·86	3·955 890	·163 058	·162 972	·163 227	− 86	+169
·90	9·129 442	·358 102	·356 913	·358 325	−1189	+223

x	y	n	$n_{(i)}$	$n_{(i)'}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(i)'}$	Error × 10^6 in $n_{(ii)}$
·94	9·540 004	·359 576	·363 964	·366 141	·358 145	+4388	+ 6561	−1431
·98	·870 514	·060 242	·050 629	·016 770	·064 435	−9613	−43472	+4193

TABLE IIᵇ.

$h = \cdot 02 = \frac{1}{2}\sigma$ approx.

x	y	n	$n_{(i)}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(ii)}$
·77	·142 978	·002 949	·002 949	·002 948	0	− 1
·79	·340 237	·006 984	·006 984	·006 983	0	− 1
·81	·757 005	·015 463	·015 463	·015 462	0	− 1
·83	1·564 403	·031 793	·031 794	·031 793	+ 1	0
·85	2·973 556	·060 106 (5)	·060 107	·060 109	+ 1	+ 3
·87	5·122 762	·102 952	·102 951	·102 960	− 1	+ 8
·89	7·817 987	·156 135	·156 127	·156 145	− 8	+10
·91	10·180 319	·201 967	·201 953	·201 968	−14	+ 1
·93	10·577 314	·208 610	·208 615	·208 580	+ 5	−30
·95	7·645 695	·150 965	·151 050	·150 915	+85	−50

x	y	n	$n_{(i)}$	$n_{(i)'}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(i)'}$	Error × 10^6 in $n_{(ii)}$
·97	2·693 587	·056 103	·056 075	·055 647	·056 148	− 28	− 456	+45
·99	·088 571	·004 139	·004 022	·009 736	·004 237	−117	+5597	+98

TABLE II^c.

Equation of Curve: $y = \dfrac{1}{\cdot 000\,000\,069\,694\,06} x^{48}(1-x)^4$.

$h = \cdot 01 = \frac{1}{4}\sigma$ approx.

x	y	n	$n_{(i)}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(ii)}$
·885	7·126 459	·071 269	·071 269	·017 269	0	0
·895	8·493 084	·084 867	·084 866	·084 867	−1	0
·905	9·701 333	·096 859	·096 859	·096 859	0	0
·915	10·536 408	·105 109	·105 109	·105 109	0	0
·925	10·760 959	·107 268	·107 267	·107 267	−1	−1
·935	10·172 193	·101 343	·101 343	·101 342	0	−1
·945	8·689 232	·086 570	·086 570	·086 568	0	−2
·955	6·453 829	·064 395	·064 396	·064 394	+1	−1
·965	3·893 916	·039 077	·039 078	·039 077	+1	0
·975	1·662 589	·017 026	·017 025	·017 029	−1	+3

x	y	n	$n_{(i)}$	$n_{(i)'}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(i)'}$	Error × 10^6 in $n_{(ii)}$
·985	·351 647	·003 946	·003 939	·003 935	·003 951	− 7	−11	+5
·995	·007 050	·000 192	·000 202	·000 281	·000 189	+10	+89	−3

TABLE III^a.

Equation of Curve: $y = \frac{1}{6}x^3e^{-x}$.

$h = 1 = \frac{1}{2}\sigma$.

x	y	n	$n_{(i)}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(ii)}$
11·5	·002 568	·002 624	·002 624	·002 624	0	0
10·5	·005 313	·005 420	·005 420	·005 420	0	0
9·5	·010 696	·010 890	·010 891	·010 890	+ 1	0
8·5	·020 826	·021 154	·021 154	·021 153	0	− 1
7·5	·038 889	·039 385	·039 387	·039 386	+ 2	+ 1
6·5	·068 814	·069 438 (5)	·069 440	·069 441	+ 2	+ 3
5·5	·113 323	·113 822	·113 822	·113 830	0	+ 8
4·5	·168 718	·168 444	·168 434	·168 457	−10	+13
3·5	·215 785	·213 762	·213 729	·213 767	−33	+ 5
2·5	·213 763	·209 892	·209 879	·209 844	− 13	−48

x	y	n	$n_{(i)}$	$n_{(i)'}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(i)'}$	Error × 10^6 in $n_{(ii)}$
1·5	·125 510	·123 888	·124 298	·124 193	·123 767	+ 410	+305	− 121
·5	·012 636	·018 988	·017 440	·018 000	·019 481	− 1548	− 988	+493

TABLE III^b.

Equation of Curve: $y = \frac{1}{6}x^3e^{-x}$.

$h = \frac{2}{3} = \frac{1}{3}\sigma$.

					Error × 10^6	
x	y	n	$n_{(i)}$	$n_{(ii)}$	in $n_{(i)}$	in $n_{(ii)}$
$7\cdot\dot{6}$	·035 162	·023 580	·023 580	·023 580	0	0
7·0	·052 129	·034 923	·034 923	·034 924	0	+ 1
$6\cdot\dot{3}$	·075 199	·050 320	·050 320	·050 321	0	+ 1
$5\cdot\dot{6}$	·104 [illegible]13	·070 107	·070 107	·070 108	0	+ 1
5·0	·140 374	·093 650 (5)	·093 650	·093 652	0	+ 2
$4\cdot\dot{3}$	·177 980	·118 509	·118 508	·118 511	− 1	+ 2
$3\cdot\dot{6}$	·210 015	·139 516	·139 514	·139 517	− 2	+ 1
3·0	·224 042	·148 441	·148 438	·148 439	− 3	− 2
$2\cdot\dot{3}$	·205 317	·135 697	·135 696	·135 688	− 1	− 9
$1\cdot\dot{6}$	·145 737	·096 382	·096 396	·096 366	+14	−16

						Error × 10^6		
x	y	n	$n_{(i)}$	$n_{(i)'}$	$n_{(ii)}$	in $n_{(i)}$	in $n_{(i)'}$	in $n_{(ii)}$
1·0	·061 313	·041 636	·041 694	·041 569	·041 632	+ 58	− 67	− 4
$0\cdot\dot{3}$	·004 423	·004 859	·004 550	·006 107	·004 969	−309	+1248	+110

TABLE III^c.

$h = \frac{1}{3} = \frac{1}{6}\sigma$.

					Error × 10^6	
x	y	n	$n_{(i)}$	$n_{(ii)}$	in $n_{(i)}$	in $n_{(ii)}$
3·5	·215 785	·071 854	·071 854	·071 854	0	0
$3\cdot1\dot{6}$	·223 044	·074 246	·074 246	·074 246	0	0
$2\cdot8\dot{3}$	·222 967	·074 195	·074 195	·074 195	0	0
2·5	·213 763	·071 109 (5)	·071 109	·071 109	0	0
$2\cdot1\dot{6}$	·194 202	·064 587	·064 587	·064 587	0	0
$1\cdot8\dot{3}$	·164 197	·054 609	·054 609	·054 609	0	0
1·5	·125 510	·041 773	·041 773	·041 772	0	−1
$1\cdot1\dot{6}$	·082 416	·027 506	·027 507	·027 506	+1	0
$\cdot8\dot{3}$	·041 917	·014 129	·014 131	·014 130	+2	+1

						Error × 10^6		
x	y	n	$n_{(i)}$	$n_{(i)'}$	$n_{(ii)}$	in $n_{(i)}$	in $n_{(i)'}$	in $n_{(ii)}$
·5	·012 636	·004 469	·004 464	·004 455	·004 466	− 5	− 14	− 3
$\cdot1\dot{6}$	·000 653	·000 390	·000 380	·000 510	·000 400	−10	+120	+10

TABLE IV.

Equation of Curve: $y = \frac{1}{8!} x^8 e^{-x}$.

$h = 1 \cdot 5 = \frac{1}{2}\sigma$.

x	y	n	$n_{(i)}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(ii)}$
17·25	·006 269	·009 634	·009 635	·009 634	+ 1	0
15·75	·013 570	·020 756	·020 757	·020 756	+ 1	0
14·25	·027 309	·041 549	·041 551	·041 551	+ 2	+ 2
12·75	·050 270	·076 032	·076 032	·076 038	0	+ 6
11·25	·082 773	·124 385	·124 380	·124 395	− 5	+10
9·75	·118 071	·176 240	·176 228	·176 245	−12	+ 5
8·25	·139 056	·206 314 (5)	·206 318	·206 304	+ 4	−10
6·75	·125 148	·185 270	·185 308	·185 235	+38	−35
5·25	·075 112	·112 505	·112 528	·112 500	+23	− 5
3·75	·022 810	·036 454	·036 366	·036 510	−88	+56

x	y	n	$n_{(i)}$	$n_{(i)'}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(i)'}$	Error × 10^6 in $n_{(ii)}$
2·25	·001 717	·003 780	·003 812	·003 987	·003 771	+32	+ 207	−9
·75	·000 001	·000 023	·000 038	− ·002 158	·000 015	+15	−2181	−8

TABLE V.

Equation of Curve: $y = \frac{1}{15!} x^{15} e^{-}$

$h = 2 = \frac{1}{2}\sigma$.

x	y	n	$n_{(i)}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(ii)}$
23	·020 924	·042 491 (5)	·042 493	·042 494	+ 2	+ 3
21	·039 501	·079 621 (5)	·079 620	·079 629	− 1	+ 8
19	·065 044	·130 140	·130 132	·130 149	− 8	+ 9
17	·090 621	·180 092	·180 084	·180 092	− 8	0
15	·102 436	·202 615	·202 626	·202 595	+11	−20
13	·088 475	·175 056	·175 083	·175 031	+27	−25
11	·053 352	·106 844	·106 838	·106 851	− 6	+ 7
9	·019 431	·040 509	·040 471	·040 541	−38	+32
7	·003 311	·007 722	·007 732	·007 725	+10	+ 3
5	·000 157	·000 504	·000 522	·000 493	+18	− 11

x	y	n	$n_{(i)}$	$n_{(i)'}$	$n_{(ii)}$	Error × 10^6 in $n_{(i)}$	Error × 10^6 in $n_{(i)'}$	Error × 10^6 in $n_{(ii)}$
3	·000 000 (5)	·000 005	− ·000 002	− ·000 028	·000 003	−7	− 33	−2
1	·000 000	·000 000	− ·000 001	·000 328	·000 000	−1	+328	0

TABLE VI.

Equation of Curve: $y = \frac{1}{24!} x^{24} e^{-x}$.

$h = 2{\cdot}5 = \frac{1}{2}\sigma$.

					Error × 10^6	
x	y	n	$n_{(i)}$	$n_{(ii)}$	in $n_{(i)}$	in $n_{(ii)}$
6·25	·000 0000	·000 000(4)	– ·000 001	·000 000(3)	– 1	0
8·75	·000 0104	·000 046(5)	·000 044	·000 044	– 2	– 2
11·25	·000 3928	·001 145(5)	·001 157	·001 138	+12	– 7
13·75	·003 5893	·009 972	·009 979	·009 975	+ 7	+ 3
16·25	·016 2364	·042 011(5)	·041 988	·042 035	–23	+24
18·75	·041 3336	·103 596	·103 585	·103 607	–11	+11
21·25	·068 4148	·169 434	·169 453	·169 416	+19	–18
23·75	·081 0450	·200 395	·200 409	·200 373	+14	–22
26·25	·073 4796	·182 388	·182 384	·182 384	– 4	– 4
28·75	·053 5346	·133 768	·133 759	·133 776	– 9	+ 8
31·25	·032 5081	·081 881	·081 878	·081 889	– 3	+ 8
33·75	·016 9210	·042 986(5)	·042 987	·042 990	+ 1	+ 4
36·25	·007 7181	·019 777	·019 779	·019 777	+ 2	0
38·75	·003 1398	·008 114	·008 115	·008 113	+ 1	– 1

TABLE VII.

Equation of Curve: $y = \frac{1}{48!} x^{48} e^{-x}$.

$h = 3{\cdot}5 = \frac{1}{2}\sigma$.

					Error × 10^6	
x	y	n	$n_{(i)}$	$n_{(ii)}$	in $n_{(i)}$	in $n_{(ii)}$
22·75	·000 0014	·000 008	·000 007	·000 008	– 1	0
26·25	·000 0421	·000 196	·000 196	·000 194	0	– 2
29·75	·000 5167	·002 111	·002 118	·002 106	+ 7	– 5
33·25	·003 2493	·012 261	·012 266	·012 263	+ 5	+ 2
36·75	·011 9707	·043 125	·043 110	·043 141	–15	+16
40·25	·028 4774	·100 038	·100 024	·100 050	–14	+12
43·75	·047 0604	·163 408	·163 419	·163 396	+11	–12
47·25	·057 1429	·197 853	·197 869	·197 831	+16	–22
50·75	·053 2811	·184 998	·184 998	·184 989	0	– 9
54·25	·039 5185	·138 094(5)	·138 085	·138 100	– 9	+ 6
57·75	·023 9927	·084 570	·084 565	·084 579	– 5	+ 9
61·25	·012 2080	·043 465	·043 466	·043 470	+ 1	+ 5
64·75	·005 3092	·019 108	·019 110	·019 108	+ 2	0
68·25	·002 0064	·007 302	·007 303	·007 301	+ 1	– 1

(ii) A Simple Non-Normal Correlation Surface.

BY H. L. RIETZ (University of Iowa).

Let x_1, x_2, x_3 be an observed set of variates each taken at random from a continuous rectangular distribution from 0 to 1. Form the sums

$$x = x_1 + x_2,$$
$$y = x_1 + x_3$$

with x_1 in common.

Assume that pairs (x, y) are thus formed, and plotted with a rectangular coordinate system. We set the problem to determine the form of the correlation surface

$$z = f(x, y),$$

where $z\,dx\,dy$ gives, to within infinitesimals of higher order, the probability that, in a trial, a point (x, y) will fall into the rectangular element $dx\,dy$.

It is known* that the theoretical distribution function, say $f_1(x)$, of marginal totals of the x-arrays of y's would be the equal sides of an isosceles triangle whose equations may be written

$$f_1(x) = x \text{ for values of } x \text{ from 0 to 1} \quad \text{.................................(1)},$$

$$\text{and } f_1(x) = 2 - x \text{ for values of } x \text{ from 1 to 2} \quad \text{...........................(2)}.$$

We now propose the question: What is the geometrical form of the theoretical array of y's that corresponds to any assigned x in dx? In answering this question, we consider two cases.

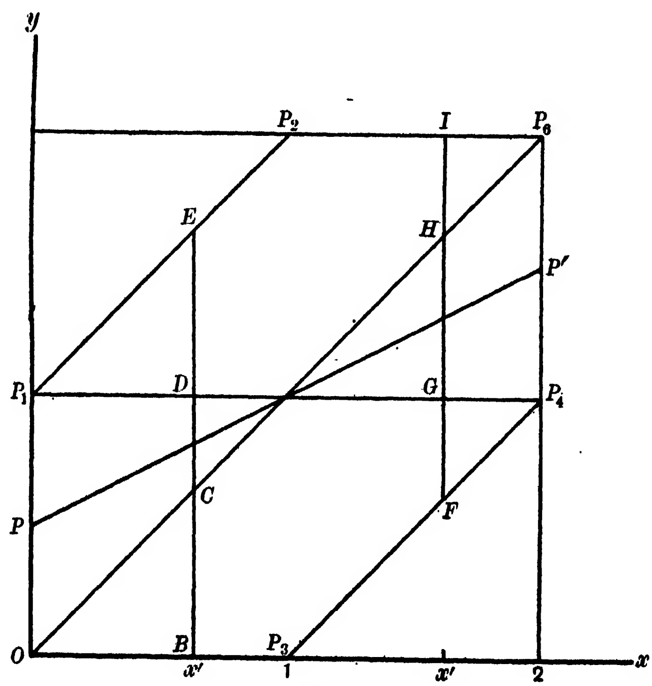


Fig. 1.

* H. L. Rietz, "Proceedings International Mathematical Congress," Vol. II. (1924) pp. 796—7; J. O. Irwin, *Biometrika*, Vol. XIX. (1927) p. 287; Philip Hall, *Biometrika*, Vol. XIX. (1927) p. 240.

Case I. *When x takes a value x' in the interval* 0 *to* 1.

It will be helpful in the exposition which follows to give attention to a square of side 2 (Fig. 1) which is the field for the scatter diagram of points (x, y).

It is obvious that no points could fall into the corner of the square cut off by the line P_1P_2 whose equation is $y=x+1$; for, $x \leqslant x' = DE$, and $x_3 \leqslant 1 = BD$.

For values of y ranging from O to x', that is, for points along BC of the x'-array of y's, the frequency of values of x_1+x_3 in an assigned dy is the same as the frequency in dy of the sum of two variates each drawn independently from a rectangular distribution from 0 to 1. Hence, the distribution curve for that part of the array of y's from B to C may be represented in the yz-plane vertical to the xz-plane of Figure 1 by the equations*

$$\left.\begin{aligned} x&=x' \\ z&=y \end{aligned}\right\} \quad\text{(3).}$$

For values of y ranging from x' to 1, that is, for points along CD, the distribution of y's is rectangular; for, no matter what value x has in the interval O to x', the chance that x_1+x_3 will fall into a given dy is the same wherever dy is taken on CD. The equation for that part of the theoretical array of y's from C to D may then be written

$$\left.\begin{aligned} x&=x' \\ z&=c \text{ where } c \text{ is constant} \end{aligned}\right\} \quad\text{(4).}$$

For values of y ranging from 1 to $1+x'$, that is, for points along DE, the distribution would again be the same as that of the sum of two independent values each taken from the interval 0 to 1. Thus, the distribution curve for the part of the array of y's from D to E is given by

$$\left.\begin{aligned} x&=x' \\ z&=-y+x'+1 \end{aligned}\right\} \quad\text{(5).}$$

Since the marginal total for the section $x=x'$ is x', we may determine c in (4) by equating the area of the total section to x'. Thus

$$\tfrac{1}{2}x'^2+c(1-x')+\tfrac{1}{2}x'^2=x'.$$

This gives $$c=x'.$$

The theoretical array of y's is an isosceles trapezoid.

Case II. *When x takes a value x' in the interval* 1 *to* 2.

When x takes a value x' in the interval 1 to 2, it follows at once that x_1 is not less than $x'-1$. Hence, it is clearly impossible for the point (x, y) to fall into the corner of the square cut off by the line P_3P_4 whose equation is $y=x-1$.

For values of y ranging from $x'-1$ to 1, that is, for points along FG, the distribution curve for the x-array of y's is given by

$$\left.\begin{aligned} x&=x' \\ z&=y-x'+1 \end{aligned}\right\} \quad\text{(6).}$$

For values of y ranging from 1 to x', that is, for points along GH, the distribution is rectangular and is given by

$$\left.\begin{aligned} x&=x' \\ z&=k \text{ where } k \text{ is a constant} \end{aligned}\right\} \quad\text{(7).}$$

Since the marginal total for $x=x'$ is $2-x'$, we determine k in (7) by equating the area of the section to $2-x'$. Thus

$$\tfrac{1}{2}(2-x')^2+k(x'-1)+\tfrac{1}{2}(2-x')^2=2-x'.$$

This gives $$k=2-x'.$$

* *loc. cit.* p. 797.

For values of y ranging from x' to 2, that is, for points along HI, the distribution curve in the x'-array of y's is given by

$$\left.\begin{array}{l} x=x' \\ z=2-y \end{array}\right\} \qquad (8).$$

It is now fairly obvious that the typical x-array of y's is an isosceles trapezoid which degenerates into straight lines when $x'=0$ and when $x'=2$, and becomes an isosceles triangle when $x'=1$.

From equations (3) to (8) we have the answer to the question about the probability that a y corresponding to an assigned x will fall into an assigned dy.

From simple geometrical considerations it follows that sections $y=y'$ of our surface $z=f(x, y)$ are similar to sections $x=x'$. The theoretical distribution of marginal totals of the y-arrays of x's is given by

$$f_2(y)= \quad y \quad \text{for } 0 \leqslant y \leqslant 1,$$

$$f_2(y)= -y+2 \text{ for } 1 \leqslant y \leqslant 2.$$

The line of regression of y's on x would be

$$y-1=\tfrac{1}{2}(x-1) \qquad (9),$$

since from symmetry the centroids of sections $x=x'$ lie on the line PP' whose slope is $\frac{1}{2}$.

Likewise, the regression line of x's on y is

$$x-1=\tfrac{1}{2}(y-1).$$

To summarise, we may describe the correlation surface by saying that for any assigned x in the interval 0 to 1, the distribution function of the array of y's is given by

$$\begin{array}{ll} z=y & \text{when } 0 \leqslant y \leqslant x, \\ z=x & \text{when } x \leqslant y \leqslant 1, \\ z=-y+x+1 & \text{when } 1 \leqslant y \leqslant 1+x, \end{array}$$

and for any assigned x in the interval 1 to 2, the distribution function of the array of y's is given by

$$\begin{array}{ll} z=y-x+1 & \text{when } x-1 \leqslant y \leqslant 1, \\ z=2-x & \text{when } 1 \leqslant y \leqslant x, \\ z=2-y & \text{when } x \leqslant y \leqslant 2. \end{array}$$

By making $x=1$ in $z=x$ or in $z=2-x$, we get $z=1$, the largest value of z. It now becomes fairly obvious that our correlation surface is simply the hexagonal pyramid of unit altitude on the base $OP_1P_2P_6P_4P_3$ with the foot of the altitude at the centre of the base.

For urn schemata with situations analogous to those of the present problem but concerned with discrete rather than with continuous variables, we may cite a paper* that considers the correlation of the sums of first and second throws with two dice.

(iii) Professor Rietz's Problem.

EDITORIAL.

This problem is more or less familiar in a more generalised form, and has been used not infrequently to explain "correlation by a common factor."

Let $y_1=x_1+x_3$, and $y_2=x_2+x_4$.

* H. L. Rietz, *Annals of Mathematics*, Vol. XXII. (1919—21) pp. 306—322.

Let σ_s be the standard deviation of x_s and Σ_s of y_s; then if r_{st} be the correlation coefficient of x_s and x_t and R of y_1, y_2, we have at once from the definitions:

$$\Sigma_1^2 = \sigma_1^2 + \sigma_3^2 + 2\sigma_1\sigma_3 r_{13} \quad \text{(i)},$$

$$\Sigma_2^2 = \sigma_2^2 + \sigma_4^2 + 2\sigma_2\sigma_4 r_{24} \quad \text{(ii)},$$

$$R = \frac{\sigma_1\sigma_2 r_{12} + \sigma_1\sigma_4 r_{14} + \sigma_2\sigma_3 r_{23} + \sigma_3\sigma_4 r_{34}}{\Sigma_1\Sigma_2} \quad \text{(iii)}.$$

Still more general equations will be found in the *Phil. Trans.* Vol. 192[A], pp. 260—261.

Professor Rietz takes his x_1 and x_2 identical and supposes x_1, x_3, x_4 drawn from the same frequency distribution. Hence

$$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$$

and

$$R = \frac{1 + r_{14} + r_{23} + r_{34}}{2\sqrt{1 + r_{13}}\sqrt{1 + r_{14}}}.$$

Supposing x_1, x_3, x_4 to be independent drawings, he has all his coefficients of correlation zero, and finds $R = \frac{1}{2}$.

The advantages of the more general form are that R can take any numerical value, x_1, x_2, x_3, x_4 can have any degree of interdependence, and they may be drawn from quite different frequency distributions.

If we suppose x_2 identical with x_1 and x_3 drawn from the same frequency distribution as x_4, we have, supposing no correlation,

$$R = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_3^2},$$

which can by properly choosing the ratio of the variances take any desired value. For example, if we wish to obtain experimental data with a correlation

$$R = n/m, \qquad m > n,$$

we toss m coins and count the heads for our first variate, we leave n on the table (a bed is better!) and toss again the $m-n$ picked up, the heads on the total m now on the table give our second variate, and the process being often repeated the correlation of the two variates will have the requisite value n/m. The experiment is of value in teaching a student how a "common factor" can be made to give any arbitrary correlation. A similar experiment with dice was habitually used by the late Professor Weldon to introduce his students to the subject of correlation.

If we suppose our pairs are drawn from a fourfold normal distribution, which in some respects is of more practical value than a rectangular one, the surface will be of the form

$$z = z_0 \text{ expt.} - \tfrac{1}{2} \text{ (quadratic function of } x_1, x_2, x_3, x_4), \quad [dx_1 dx_2 dx_3 dx_4].$$

Now replace x_1 by $y_1 - x_3$ and x_2 by $y_2 - x_4$, and we have

$$z = z_0 \text{ expt.} - \tfrac{1}{2} \text{ (quadratic function of } y_1, y_2, x_3, x_4), \quad [dy_1 dy_2 dx_3 dx_4].$$

Integrate out x_3 and x_4 between the limits $\pm\infty$, and we have

$$z = z_0' \text{ expt.} - \tfrac{1}{2} \text{ (quadratic function of } y_1, y_2), \quad [dy_1 dy_2].$$

We need not, however, perform these operations in order to determine what the coefficients of y_1^2, y_1y_2 and y_2^2 are*. They will be given by

$$\frac{N}{2\pi\Sigma_1\Sigma_2\sqrt{1-R\Sigma}} e^{-\frac{1}{2}\frac{1}{1-R^2}\left(\frac{y_1^2}{\Sigma_1^2} - \frac{2Ry_1y_2}{\Sigma_1\Sigma_2} + \frac{y_2^2}{\Sigma_2^2}\right)}, \quad [dy_1 dy_2] \quad \text{(iv)},$$

where Σ_1, Σ_2 and R are provided by Equations (i)—(iii) above.

Whatever the distribution may be, if x_1 $(=x_2)$, x_3 and x_4 are drawn from that distribution only, $R = \frac{1}{2}$.

* I once did so; it forms an interesting exercise in determinantal analysis, and of course concludes with (iv).

(iv) Note on a Memoir by A. E. R. Church.

Biometrika, Vol. XVIII.

In a memoir on the "Means and Squared Standard Deviations of Small Samples" of 1926 I gave, p. 382, a formula for the fourth moment coefficient of the variance of samples of N from a limited population of size M following any law of frequency. A printer's error was discovered, namely the number 217 was read for 271, in the ninth line of p. 382, and this was published in *Biometrika*, Vol. XIX, list of *Errata* at the beginning. Lately, in a memoir published by Dr L. Isserlis in the *R.S. Proc.* A, Vol. 132, 1931, further errors are referred to, namely p. 382, line 5, $-18N+77$ is read for $-42N+133$; line 6, $+2880$ for -2880, and line 7, $+21840$ for -5040. Having verified these corrections I am very grateful to Dr Isserlis for drawing my attention to them and hope those possessing copies of *Biometrika* will note the changes mentioned.

As I stated in my memoir, the practical use of a formula covering three and a half pages of *Biometrika* is very small; in fact the main object of its publication was to indicate that the moment coefficients of moment coefficients of samples from finite populations have reached, with the variance, unmanageable dimensions, and that general discussions on such moment coefficients, though they may be excellent examples of algebraic manipulation, are of small statistical value, for the results are quite unusable in practice.

BIOMETRIKA

FURTHER APPLICATIONS IN STATISTICS OF THE $T_m(x)$ BESSEL FUNCTION.

BY KARL PEARSON, S. A. STOUFFER AND F. N. DAVID*.

(1) THE $T_m(x)$ function was defined in a paper by Pearson, Jeffery and Elderton† to be given by

$$T_m(x)=\frac{1}{\sqrt{\pi}}\frac{1}{2^m}\frac{1}{\Gamma(m+\frac{1}{2})}x^m K_m(x)\qquad\text{(i)},$$

where $K_m(x)$ is the Bessel Function of the second order and imaginary argument. Here $T_m(x)=T_m(-x)$, while x on the right is always to be given its numerical value. Remembering this, we need not write $|x|^m K_m(|x|)$ in the equation.

If

$$y=MT_m(x)\qquad\text{(ii)}$$

be treated as a frequency curve, it will be symmetrical and run from $-\infty$ to $+\infty$ of x. The constant in (i) has been so chosen that

$$\int_{-\infty}^{+\infty} y\,dx=2M\int_0^\infty T_m(x)\,dx=M.$$

An integral form of $K_m(x)$ is given by‡

$$K_m(x)=\frac{\sqrt{\pi}\,x^m}{2^m\,\Gamma(m+\frac{1}{2})}\int_1^\infty e^{-xt}(t^2-1)^{m-\frac{1}{2}}\,dt\qquad\text{(iii)}.$$

Hence we may write (ii) in the form

$$y=\frac{M}{2^{2m}}\frac{1}{\Gamma^2(m+\frac{1}{2})}x^{2m}\int_1^\infty e^{-xt}(t^2-1)^{m-\frac{1}{2}}\,dt\qquad\text{(iv)}.$$

(2) Consider in the next place the curve

$$y=y_0 e^{-\frac{px}{a}}\left(1+\frac{x}{a}\right)^p\qquad\text{(v)},$$

the origin being the mode at distance a from the start of the curve.

It follows easily that

$$y_0=\frac{M}{a}\frac{p^{p+1}e^{-p}}{\Gamma(p+1)}\qquad\text{(vi)},$$

where M is the total frequency.

* The suggestion of the problem and the selection of the illustrative examples were provided by S. A. Stouffer, the solution through the $T_m(x)$ function was given by K. Pearson, who is also responsible for the text. Florence N. David computed the table of the probability integral of the $T_m(x)$ distribution.

† *Biometrika*, Vol. xxi. p. 184.

‡ G. N. Watson: *A Treatise on the Theory of the Bessel Functions*, p. 172, Equation (4).

Thus the curve can be written

$$y = M\frac{p}{a}\frac{e^{-p\left(1+\frac{x}{a}\right)}}{\Gamma(p+1)}\left\{p\left(1+\frac{x}{a}\right)\right\}^p \text{ ....(vii).}$$

Write $z = p\left(1+\frac{x}{a}\right)$ and the moments about the start of the curve can be found at once. These lead to*

$$\left.\begin{aligned} \text{Mean} = \bar{x}' &= a(p+1)/p \\ \text{Standard Deviation} = \sigma &= a\sqrt{p+1}/p \\ \beta_1 = \frac{4}{p+1}, \quad \beta_2 &= 3 + \frac{6}{p+1} \end{aligned}\right\} \text{ ....(viii),}$$

providing the well-known relation, $2\beta_2 - 3\beta_1 - 6 = 0$.

(3) Now suppose there are two independent variates u and v both of which have frequency distributions provided by Equation (vii). We assume the two distributions to have the same p, but to have different standard deviations σ_1 and σ_2, or, what amounts to the same thing, different modal distances a and b. We will measure our variates u and v from the start of their curves, which then take the form

$$y_1 = M\frac{p}{a}e^{-\frac{pu}{a}}\left(\frac{pu}{a}\right)^p \Big/ \Gamma(p+1),$$

and

$$y_2 = M\frac{p}{b}e^{-\frac{pv}{b}}\left(\frac{pv}{b}\right)^p \Big/ \Gamma(p+1).$$

If we take $w = M\frac{y_1}{M}\times\frac{y_2}{M}$, we obtain the combined frequency surface

$$w = M\frac{p}{a}\frac{p}{b}\frac{1}{\Gamma^2(p+1)}e^{-\left(\frac{pu}{a}+\frac{pv}{b}\right)}\left(\frac{pu}{a}\frac{pv}{b}\right)^p \text{ ....(ix).}$$

Now put $X = p\left(\frac{u}{a}+\frac{v}{b}\right)$ and $Y = p\left(\frac{v}{b}-\frac{u}{a}\right)$, then the element for integration of the above surface is $du\,dv$, or if we take it $d\left(\frac{pu}{a}\right)d\left(\frac{pv}{b}\right)$ we may replace it by $dX\,dY$, and we have for integration

$$\frac{M}{\Gamma^2(p+1)\,2^{2p}}e^{-X}(X^2 - Y^2)^p\,dX\,dY \text{ ....(x).}$$

We have to integrate this out for X to get the distribution curve of Y. In the upper octant XOB (Fig. 1, p. 295) the limit for X is clearly $X = Y$ to $X = \infty$ along the shaded area. Or, the curve of distribution of Y is

$$z = \frac{M}{\Gamma^2(p+1)\,2^{2p}}\int_Y^X e^{-X}(X^2 - Y^2)^p\,dX \text{ ....(x \textit{bis}).}$$

Put $X = Yt$ and we have

$$\frac{M}{\Gamma^2(p+1)\,2^{2p}}Y^{2p+1}\int_1^\infty e^{-Yt}(t^2-1)^p\,dt \text{ ....(xi).}$$

* *Phil. Trans.*, Vol. 185A, p. 373.

If we take the lower octant XOA, the limits of X are $-Y$ to ∞, but as Y is now negative we get precisely the same result, or we say that the whole curve of distribution of Y is (xi), Y being taken as positive, and from 0 to ∞, and mirrored in the axis of X. This result also flows from the fact that the distribution of $\frac{v}{b}-\frac{u}{a}$ must be a symmetrical curve, as the frequency curves for u/a and v/b are identical.

Now if in (iv) we write $x=Y$, $m=p+\frac{1}{2}$, we see that the z of (xi) is given by

$$z=MT_{p+\frac{1}{2}}(Y) \quad\text{.................................(xii)},$$

which leads to $\frac{1}{2}M$ for the area of our half curve. In other words our curve for Y is the $T_{p+\frac{1}{2}}$ curve mirrored on itself. The ordinates of this curve have been computed by Dr E. M. Elderton*.

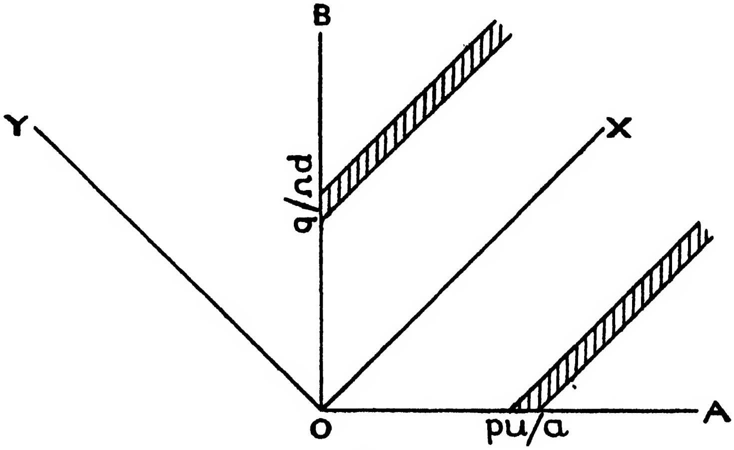


Fig. 1.

(4) Now the odd moments of the mirrored curve vanish. Let us find the even moment-coefficients. We have from (x)

$$M\mu_{2s}=2\frac{M}{\Gamma^2(p+1)\,2^{2p}}\iint Y^{2s}e^{-X}(X^2-Y^2)^p\,dY\,dX,$$

where the limits of X and Y are to be chosen so as to cover the upper octant BOX. Now if we integrate first with regard to Y, the limits will be from 0 to X, and then with regard to X from 0 to ∞. Thus

$$\mu_{2s}=\frac{1}{2^{2p-1}\Gamma^2(p+1)}\int_0^\infty e^{-X}\int_0^X Y^{2s}(X^2-Y^2)^p\,dY\,dX \quad\text{.........(xiii)}.$$

Put $Y=X\lambda$ and we have

$$\mu_{2s}=\frac{1}{2^{2p-1}\Gamma^2(p+1)}\int_0^\infty e^{-X}X^{2s+2p+1}\int_0^1\lambda^{2s}(1-\lambda^2)^p\,d\lambda\,dX,$$

* See *Biometrika*, Vol. XXI, pp. 194—201, or *Tables for Statisticians and Biometricians*, Part II. pp. lxxix—lxxxviii and 138—144.

or, if $\lambda^2 = \kappa$,

$$\mu_{2s} = \frac{1}{2^{2p}\Gamma^2(p+1)}\Gamma(2s+2p+2)\int_0^1 \kappa^{s-\frac{1}{2}}(1-\kappa)^p\,d\kappa$$

$$= \frac{1}{2^{2p}\Gamma^2(p+1)}\Gamma(2s+2p+2)\frac{\Gamma(s+\frac{1}{2})\Gamma(p+1)}{\Gamma(s+p+\frac{3}{2})}.$$

If $s = 0$,
$$\mu_0 = \frac{1}{2^{2p}\Gamma(p+1)}\,\frac{\Gamma(2p+2)\Gamma(\frac{1}{2})}{\Gamma(p+\frac{3}{2})} = 1.$$

Hence
$$\mu_{2s} = \frac{\Gamma(2s+2p+2)}{\Gamma(2p+2)}\,\frac{\Gamma(p+\frac{3}{2})}{\Gamma(s+p+\frac{3}{2})}\,\frac{\Gamma(s+\frac{1}{2})}{\Gamma(\frac{1}{2})} \quad\text{(xiv)},$$

and
$$\mu_2 = \frac{(2p+3)(2p+2)}{p+\frac{3}{2}}\,\frac{1}{2} = 2p+2 \quad\text{(xv)}.$$

Generally
$$\mu_{2s} = (2s-1)(2p+2s)\,\mu_{2s-2} \quad\text{(xv } bis\text{)},$$

$$\beta_{2s-2} = \frac{\mu_{2s}}{(\mu_2)^s} = \frac{(2s-1)(2p+2s)}{2p+2}\,\frac{\mu_{2s-2}}{(\mu_2)^{s-1}},$$

or,
$$\beta_{2s-2} = (2s-1)\left(1+\frac{2(s-1)}{2(p+1)}\right)\beta_{2s-4} \quad\text{(xvi)}.$$

Thus finally

$$\beta_{2s-2} = (2s-1)(2s-3)\ldots 1\left(1+\frac{s-1}{p+1}\right)\left(1+\frac{s-2}{p+1}\right)\ldots\left(1+\frac{1}{p+1}\right)\ldots\text{(xvii)}.$$

It will be clear that when $p \to \infty$ we obtain

$$\beta_{2s-2} = (2s-1)(2s-3)\ldots 1,$$

the familiar β_{2s-2} formula for the normal curve, into which the $T_{p+\frac{1}{2}}$ function then passes.

Consider the Type VII curve

$$y = y_0\frac{1}{(a^2+x^2)^{\frac{1}{2}n}}.$$

Here we have

$$\beta_{2s-2} = (2s-1)(2s-3)\ldots 1\left(1+\frac{2(s-1)}{n-2s-1}\right)\left(1+\frac{2(s-2)}{n-2s+1}\right)\ldots\left(1+\frac{2}{n-5}\right)$$

and $\mu_2 = a^2/(n-3)$.

Now it is clear that we can make μ_2 and μ_4 agree in the Type VII and the $T_{p+\frac{1}{2}}$ curves*, but farther than that we cannot go, although the β_4's may not differ widely if n be considerable. The $T_{p+\frac{1}{2}}$ curve has the further advantage that no moment-coefficients tend to become infinite, while if n be an odd integer, those for the Type VII curve may become so. For values of p not too great the Type VII will fit the distribution of Y considerably better than the normal curve. For considerable values of p, both Type VII and the $T_{p+\frac{1}{2}}$ curves pass into the normal curve.

(5) A few further points may be noted. If $p = -\frac{1}{2}$ the T_0-curve asymptotes to the vertical at the origin, and this holds as long as p lies between $-\frac{1}{2}$ and 0; if

* We must take $\frac{1}{p+1} = \frac{2}{n-5}$ or $n = 2p+7$, and $a = 2\sqrt{(p+1)(p+2)}$.

$p=0$, the $T_{\frac{1}{2}}$-curve starts with a finite ordinate and makes a finite angle with the vertical, it is the exponential curve. If p be positive we see from (x *bis*) that $dz/dY=0$ for $Y=0$, or the double mirror curves have a common tangent at the axis of symmetry and will *in appearance* form a single curve. If p be a positive integer it is possible to expand z in powers of Y, but the series does not present any great advantages to the computer.

When $p=11$, Dr Elderton's Tables terminate, but it is shown in the memoir by Pearson, Jeffery and Elderton* that when $p=11$, the two curves

$$z=MT_{p+\frac{1}{2}}(Y)$$

and

$$z=\frac{M}{\sqrt{2\pi(p+1)(p+2)}}\,\frac{\Gamma\{\frac{1}{2}(2p+7)\}}{\Gamma(p+3)}\,\frac{1}{\left(1+\dfrac{Y^2}{4(p+1)(p+2)}\right)^{\frac{1}{2}(2p+7)}}\quad\text{......(xviii)}$$

coincide for practical statistical purposes. The areas of this latter curve up to given values of Y have been tabled† from $p=-\frac{1}{2}$ to $p=12$, but this hardly carries us beyond the T_m-tables. The completed (and now at press) *Tables of the Incomplete B-function* carry us up to $2p+7=101$, or $p=47$.

(6) Now let us turn to the means of samples of size n drawn from the Type III curve

$$y=y_0'\,e^{-\frac{px}{a}}\left(\frac{x}{a}\right)^p\quad\text{..............................(xix)},$$

where the origin is at the start of the curve and a is the distance to the mode from the start. Let us suppose a sample $x_1, x_2, x_3 \ldots x_n$ drawn and let its mean be $\bar{x}_n=(x_1+x_2+\ldots+x_n)/n$. Then the chance P of a sample lying between x_1 and $x_1+\delta x_1$, x_2 and $x_2+\delta x_2$, ... x_n and $x_n+\delta x_n$ is given by

$$P=\text{const.}\times e^{-\frac{p}{a}(x_1+x_2+\ldots+x_n)}\left(\frac{x_1x_2\ldots x_n}{a^n}\right)^p dx_1dx_2\ldots dx_n.$$

Now get rid of x_1 by introducing $\bar{x}_n$ as a variable and write l_2 for

$$n\bar{x}_n-x_3-x_4-\ldots-x_n.$$

We have

$$P=\text{const.}\times e^{-\frac{np\bar{x}_n}{a}}\,d\bar{x}_n\left(\frac{l_2-x_2}{a}\right)^p\left(\frac{x_2}{a}\right)^p\left(\frac{x_3\ldots x_n}{a^{n-2}}\right)^p dx_2dx_3\ldots dx_n.$$

Put $x_2=l_2x_2'$ and integrate out for $x_2=0$ to l_2 or $x_2'=0$ to 1. This will introduce a B-function into the constant, but leave us with

* Cf. *Biometrika*, Vol. XXI. pp. 171 and 173 for accordance of the curves. Their equations are given on p. 185, where we must write $\frac{1}{2}n-1=p+\frac{1}{2}$, or $n=2p+3$. The two curves have then the same first four moment-coefficients. If $\eta=Y/\{2\sqrt{(p+1)(p+2)}\}$, then the proportional area from $\eta=0$ up to any arbitrary value of η is given by $\frac{1}{2}I_\eta(\frac{1}{2}, p+1)$, where $I_\eta(\frac{1}{2}, p+1)=B_\eta(\frac{1}{2}, p+1)/B(\frac{1}{2}, p+1)$, B_η and B being the incomplete and complete Beta-functions.

† See *Biometrika*, Vol. XXII. pp. 253—283, or *Tables for Statisticians and Biometricians*, Part II, pp. cxxv—cxlii and pp. 169—177.

$$P = \text{const.} \times e^{-\frac{np\bar{x}_n}{a}} d\bar{x}_n \left(\frac{l_2}{a}\right)^{2p+1} \left(\frac{x_3 \dots x_n}{a^{n-2}}\right)^p dx_3 \dots dx_n.$$

Write $l_2 = l_3 - x_3$, and proceeding in the same way, we find

$$P = \text{const.} \times e^{-\frac{np\bar{x}_n}{a}} d\bar{x}_n \left(\frac{l_3}{a}\right)^{3p+2} \left(\frac{x_4 \dots x_n}{a^{n-3}}\right)^p dx_4 \dots dx_n,$$

where $l_3 = n\bar{x}_n - x_4 - x_5 - \dots - x_n$.

Continuing to repeat this process we ultimately get rid of all the variables but $\bar{x}_n$ and find*

$$P = \text{const.} \times e^{-\frac{np\bar{x}_n}{a}} \left(\frac{\bar{x}_n}{a}\right)^{n(p+1)-1} d\bar{x}_n \quad \text{.....................(xx).}$$

We now put this into the canonical form for a Type III frequency curve, i.e.

$$y = y_0 e^{-\frac{P}{A}\bar{x}_n} \left(\frac{\bar{x}_n}{A}\right)^P \quad \text{..........................(xx } bis\text{).}$$

Hence we must have $P = n(p+1) - 1$, and $P/A = np/a$, or $A = a\dfrac{n(p+1)-1}{np}$.

Accordingly:

$$\text{Mode of } \bar{x}_n = a\,\frac{n(p+1)-1}{np}$$

$$\left.\begin{aligned}\text{Mean of } \bar{x}_n = M_1' = \frac{A(P+1)}{P} = \frac{a(p+1)}{p} = \bar{x}\end{aligned}\right\} \quad \text{...........(xxi),}$$

$$\sigma_{\bar{x}_n}{}^2 = M_2' - \frac{A^2(P+1)}{P^2} = \frac{1}{n}\,\frac{a^2(p+1)}{p^2} = \frac{1}{n}\sigma_x{}^2$$

where $\bar{x}$ and σ_x are the mean and standard deviation of the population from which the sample of n is drawn. Lastly

$$B_1 = \frac{4}{n(p+1)} \quad \text{and} \quad B_2 = 3 + \tfrac{3}{2} B_1 \text{.....................(xxii).}$$

Clearly, if n and p are not very small, then (xx *bis*) will approach much nearer to a normal distribution than the parent population (xix).

(7) We can now apply our results to particular cases. If we draw two individuals out of Type III curves like (xix), with the same skewness as measured by p, then if a and a' be their modal distances, and

$$Y = p\left(\frac{x_2}{a_2} - \frac{x_1}{a_1}\right) = (p+1)\left(\frac{x_2}{\bar{x}_2} - \frac{x_1}{\bar{x}_1}\right) = \sqrt{(p+1)}\left(\frac{x_2}{\sigma_{x_2}} - \frac{x_1}{\sigma_{x_1}}\right),$$

for these are all equivalent, then the distribution of Y is given by

$$z = MT_{p+\frac{1}{2}}(Y).$$

If the two individuals are taken from absolutely the same population, i.e. $a_2 = a_1 = a$, then

$$Y = p\,\frac{x_2 - x_1}{a} = (p+1)\,\frac{x_2 - x_1}{\bar{x}} = \sqrt{(p+1)}\,\frac{x_2 - x_1}{\sigma_x}.$$

* This result was published by Church: see *Biometrika*, Vol. XVIII. p. 386.

Such results, however interesting in the case of experimental sampling in the Laboratory, where we have a knowledge of the parent population, will hardly be of practical service, because we should usually lack a knowledge of p, $\bar{x}$ and σ_x.

Now turn to (xx), and suppose we have taken two samples of n and that their means are $\bar{x}_n$ and $\bar{x}_n'$, then the distribution of $Y = \frac{P}{A}(\bar{x}_n' - \bar{x}_n)$ will be

$$z = \tfrac{1}{2} M T_{p+\frac{1}{2}}(Y) = \tfrac{1}{2} M T_{n(p+1)-\frac{1}{2}}(Y) \quad \text{.................(xxiii).}$$

There are now a variety of ways in which it is possible to express Y. In the first place $P/A = \frac{np}{a}$, where p and a refer to the parent population, but mean $-$ mode $= \frac{a}{p} = \bar{x} - \tilde{x}$, say. Again $\frac{p}{a} = \frac{\bar{x}}{\sigma_x^2} = \frac{2}{\sqrt{\beta_1}\,\sigma_x}$. Thus we have

$$Y = n\frac{\bar{x}_n' - \bar{x}_n}{\bar{x} - \tilde{x}} = \frac{n\bar{x}(\bar{x}_n' - \bar{x}_n)}{\sigma_x^2} = \frac{2n}{\sqrt{\beta_1}}\frac{(\bar{x}_n' - \bar{x}_n)}{\sigma_x} \quad \text{...........(xxiv).}$$

Further, we need the value of the $p+1$ in the degree of the T_m function; we have

$$p + 1 = \frac{\bar{x}}{\bar{x} - \tilde{x}} = \frac{\sigma_x^2}{(\bar{x} - \tilde{x})^2} = \frac{4}{\beta_1} \quad \text{.......................(xxv).}$$

Here $\bar{x}$, $\tilde{x}$, σ_x and β_1 all refer like p to the parent population. Clearly some two of these quantities $\bar{x}$ and $\tilde{x}$, $\bar{x}$ and σ_x, or β_1 and σ_x must be known, or we cannot determine a and p. We shall see later that in certain other applications p is known, and then probably σ_x is the best quantity to seek for. It might be thought that $\bar{x}$ would be easy to find. It may be so, if the start of the curve can be determined, but it must be remembered that $\bar{x}$ is the mean measured from a definite point of the parent population, i.e. the start of the parent population, and this may be quite unknown, $\bar{x} - \tilde{x}$ does not involve this knowledge, but the mode is not an easily determined character. On the whole β_1 and σ_x can probably be most easily obtained from the samples. Of course this refers to cases in which the parent population is unknown, but suspected of having a skewness which may be approximated to by a Type III curve The procedure here would be to determine to the second and third moment coefficients of the pooled samples, and thus obtain the best approximation which is available to β_1 and σ_x of the supposed parent population.

We then take $m = \frac{4n}{\beta_1} - \frac{1}{2}$, and

$$Y = \frac{2n}{\sqrt{\beta_1}}\frac{\bar{x}_n' - \bar{x}_n}{\sigma_x} \quad \text{.........................(xxvi),}$$

and test whether the probability integral of $T_m(Y)$ has a value sufficiently large to justify us in assuming that $\bar{x}_n'$ and $\bar{x}_n$ came from the same population.

Perhaps a more useful case occurs when one sample is sufficiently large to give reasonable values for the constants, and we ask whether the other could have been drawn from the same population. In this case we may determine p and a with sufficient accuracy from the large sample and measure the probability of x_n for the second sample from (xx) or (xx *bis*) by aid of the *Tables of the Incomplete* Γ*-function*.

Generally we may state our problem to be this: We wish to know from the means of two samples whether they are consistent with these samples having been both drawn from the same population. We have no reason for supposing that population follows a normal distribution, or we may have good reason for supposing its distribution skew. Shall we do better to assume $\beta_1 = 0$ and the unknown parent population to be normal, or to work with the value of β_1 found from the pooled samples? Probably with samples of 25 or 20 the latter would be the wiser course; at any rate, on comparison with the former method, it would give us some measure of the extent to which skewness might invalidate our conclusions from the normal hypothesis. Unfortunately we do not at present know the distribution of the variance of samples drawn from a Type III curve, or indeed from any skew curve. Had we known it, it might be possible to construct a quantity like "Student's z" with the additional advantage, however, that it would possess correlation between numerator and denominator.

(8) Type III curve gives the distribution of frequency for other statistical functions than that of the means of samples drawn from a Type III distribution. One of the most important cases is that of the distribution of the standard deviations (or of the variances) of samples from a normal parent population. If Σ be the standard deviation, M_2 the variance of the parent population $= \Sigma^2$, n the size of the sample, σ its standard deviation, μ_2 its variance, we have Helmert's Equation for the distribution of σ,

$$y = y_0 \left(\frac{\sqrt{n}\sigma}{\Sigma}\right)^{n-2} e^{-\frac{n\sigma^2}{2\Sigma^2}} \; [d\sigma] \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xxvii)},$$

or, expressed in terms of the variances, the Type III equation

$$y = y_0' \left(\frac{\sqrt{n}\mu_2}{M_2}\right)^{\frac{n-3}{2}} e^{-\frac{n\mu_2}{2M_2}} \; [d\mu_2] \;\ldots\ldots\ldots\ldots\ldots\text{(xxviii)}.$$

Hence if we have two samples of size n with variances μ_2, μ_2' taken from normal parent populations of variance M_2 and M_2', the distribution of the difference

$$Y = \frac{n}{2}\left(\frac{\mu_2'}{M_2'} - \frac{\mu_2}{M_2}\right)$$

is given by $$z = \tfrac{1}{2} M T_{\frac{1}{2}(n-2)}(Y) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xxix)}.$$

We are therefore in a position to determine whether the variances of two samples each measured in terms of the variance of its parent population are significantly different.

If the two parent populations are identical, then $Y = \frac{n}{2}(\mu_2' - \mu_2)/M_2$. If, as may often be the case, the parent population be unknown, then the only remedy is to take for M_2 the value provided by the two samples pooled. If we know the means $\bar{x}_n$ and $\bar{x}_n'$ of the two samples to be the same, this will be $\frac{1}{2}(\mu_2' + \mu_2)$, so that the frequency of the difference will be given by

$$\tfrac{1}{2} M T_{\frac{1}{2}(n-2)} \left(\frac{n(\mu_2' - \mu_2)}{\mu_2' + \mu_2}\right) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xxx)}.$$

If on the other hand we know that $\bar{x}_n$ is not equal to $\bar{x}_n'$, we have to put

$$M_2 = \tfrac{1}{2}(\mu_2' + \mu_2) + \tfrac{1}{4}(\bar{x}_n' - \bar{x}_n)^2,$$

and have accordingly

$$z = \tfrac{1}{2}MT_{\frac{1}{2}(n-2)}\left(\frac{n(\mu_2' - \mu_2)}{\mu_2' + \mu_2 + \frac{1}{2}(\bar{x}_n' - \bar{x}_n)^2}\right) \quad \text{...............(xxxi).}$$

Dr Elderton's Tables provide the ordinates* of the above curves up to samples of $n = 25$. For large samples at present we are thrown back on the Type VII curve, the probability integral of which is included in the *Incomplete* B-*function Tables.*

(9) If we have a population of large size M consisting of v categories whose frequencies are

$$m_1, m_2, \dots m_v,$$

and a sample of size N be taken giving categories of size

$$n_1, n_2, \dots n_v$$

only restricted by the condition $\overset{v}{S}(n_s) = N$, and we form the quantity

$$\chi^2 = \frac{(n_1 - \bar{n}_1)^2}{\bar{n}_1} + \frac{(n_2 - \bar{n}_2)^2}{\bar{n}_2} + \dots + \frac{(n_v - \bar{n}_v)^2}{\bar{n}_v} \quad \text{............(xxxii),}$$

where $\bar{n}_s = m_s N/M$, then the distribution of χ follows the curve†

$$y = y_0 e^{-\frac{1}{2}\chi^2}\chi^{v-2} \quad [d\chi] \text{.....................(xxxiii)}$$

and the distribution of $\frac{1}{2}\chi^2$, the Type VII curve

$$y = y_0' e^{-\frac{1}{2}\chi^2}(\tfrac{1}{2}\chi^2)^{\frac{v-3}{2}} \quad [d(\tfrac{1}{2}\chi^2)] \quad \text{.........(xxxiv).}$$

Now this Type VII curve, like that for the variance, has its p known, $= \frac{1}{2}(v-3)$, which can be found at once from v the number of categories in the sample. Further, its a, i.e. its modal distance, $= \frac{1}{2}(v-3)$, and its standard deviation $= \sqrt{\frac{1}{2}(v-1)}$. Again, if required, $\beta_1 = \dfrac{12}{v-1}$ and $\beta_2 = 3 + \dfrac{8}{v-1}$.

Thus if we have two χ^2's, namely χ^2 and χ'^2, from two samples, the distribution of their difference will be given by

$$z = \tfrac{1}{2}MT_{\frac{1}{2}(v-2)}(\tfrac{1}{2}\chi'^2 - \tfrac{1}{2}\chi^2) \quad \text{.....................(xxxv).}$$

Accordingly we have obtained a measure of whether two χ^2's supposed to be due to sampling from the same population are reasonably probable. Given the χ^2's the solution is independent of any knowledge of the parent population, beyond the number of categories used in determining the χ^2's.

We may make some remarks on the curve in (xxxv). The standard deviation of a T_m curve is $\sqrt{2m+1}$ and in our case $m = \frac{1}{2}(v-2)$; therefore, if $Y = \frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2$,

$$\sigma_Y = \sqrt{v-1} = \sqrt{\tfrac{1}{2}(v-1) + \tfrac{1}{2}(v-1)} = \sqrt{\sigma^2_{\frac{1}{2}\chi'^2} + \sigma^2_{\frac{1}{2}\chi^2}},$$

as it should do, since $\frac{1}{2}\chi'^2$ and $\frac{1}{2}\chi^2$ are by hypothesis due to independent samples.

* A probability integral table of $T_m(Y)$ accompanies this paper.

† Pearson, *Phil. Mag.* 1899, p. 239. For properties of the χ^2 curve, see *Drapers' Company Research Memoirs, Biometric Series*, VIII.

Again, $B_1 = 0$, of course, since the T_m curve is symmetrical, and $B_2 = 3 + \frac{6}{v-1}$ *. Accordingly, $B_2 - 3 = \frac{1}{2}(\beta_2 - 3)$ of the $\frac{1}{2}\chi^2$ distribution or the kurtosis is halved, i.e. the $\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2$ curve is 50°/ₒ less leptokurtic than the parent population.

Further, the distribution of the $\frac{1}{2}\chi^2$ difference does not depend *directly* † on the

* See above, p. 294, Equation (viii).

† Indirectly, of course, it does, a point too often overlooked. Take, for example, the data for Typhoid Inoculated and Attacked, due to Greenwood and Yule.

	Attacked	Not Attacked	Totals
Inoculated ...	56	6,759	6,815
Non-Inoculated	272	11,396	11,668
Totals	328	18,155	18,483

What is the exact meaning of the 6,815 and 11,668? Are they simply samples of the two classes Inoculated and Non-Inoculated, that the recorders have taken, or have they some relation to the numbers in the community willing or unwilling to be inoculated? If the former, they are subject to the more or less arbitrary choice of the recorders, and if we find χ^2 we are finding it subject to the supposition that the recorders repeatedly made experiments with the same numbers. In this case there is only one degree of freedom in this table, or $\kappa - 1$ degrees, when two populations with κ categories are compared. It seems in many respects more advantageous to treat this problem in the manner it was first investigated, namely as the comparison of two linear series, when the limitation on the degrees of freedom is seen at once to arise and arise naturally (*Biometrika*, Vol. VIII. pp. 250—254).

But suppose the numbers of Inoculated and Non-Inoculated arise from some natural division, as in the case of vaccination, and non-vaccination in the country at large, then our table represents an arbitrary sample of the total population, and there are *three* degrees of freedom, and this must be borne in mind in determining the probability of the observed result. In such a case we may or may not know the relative frequencies of the inoculated and non-inoculated in the population under consideration. If we do not, the only thing we can do is to use the observed ratio, as if it were the population ratio of the two classes. In the case of the ratio of inoculated to non-inoculated following as a natural order, i.e. a table obtained by a random sample out of a general population, where we select an individual without regard to whether he has been inoculated or attacked and afterwards inquire into details, the χ^2 is simply proportional to the total number of individuals selected. Thus the χ^2 for the above table is 56·284, but had we taken a sample of half the size it would be (subject to variation of sampling) 28·117. In other words the value of χ^2 and accordingly of P depends very largely on the size of the sample, and the comparison of χ^2's for two tables of different totals can be made to give almost any value we please to the probability of the two χ^2's being due to samples of different sizes from the same parent population. The quantity which would remain approximately the same would be the ϕ^2, and in comparing two tables like the above of *different* sizes to test whether they come from the same population, it is rather the comparison of ϕ^2 and ϕ'^2, than of χ^2 and χ'^2 which should guide us.

If, on the other hand, the sizes of the two samples of Inoculated and of Non-Inoculated in a table like that above have been arbitrarily selected, the χ^2 will change widely with those sizes. For example, if we suppose the table formed by two independent samplings, one of the Inoculated and another of the Non-Inoculated, and then recording whether they had been attacked or not, the vertical marginal tables are at our choice, and for five arbitrary sizes of the two samples we have approximately:

(*a*)

56	6,759	6,815
272	11,396	11,668
328	18,155	18,483

(*b*)

56	6,759	6,815
159	6,656	6,815
215	13,415	13,630

(*c*)

56	6,759	6,815
68	2,849	2,917
124	9,608	9,732

size of the samples on which χ^2 and χ'^2 are based, but solely on the number of cells used in computing the χ^2's being the same.

A point here is, perhaps, worth noting as we have not seen it recorded. If we take two sets of N samples each, and $m_{\bar{x}_n}$, $m_{\bar{x}_n}'$ be the means of the means of the two sets, then $m_{\bar{x}_n}$ and $m_{\bar{x}_n}'$ will be distributed according to a Type VII curve, if the original parent population is, because in this case $\bar{x}_n$ and $\bar{x}_n'$ are so distributed,

(d)

28	3,379	3,407
136	5,698	5,834
164	9,077	9,241

(e)

28	3,379	3,407
68	2,849	2,917
96	6,228	6,324

These tables give:

	(a)	(b)	(c)	(d)	(e)
χ^2	56·234	50·135	36·998	28·117	19·6802
ϕ^2	·003,042	·003,578	·003,802	·003,042	·003,112

A similar variation of χ^2 arises if we take two arbitrary samples of attacked and not attacked, and inquire as to whether they were inoculated or not. In other words, when there is no "natural" proportion of the sizes of the two samples which are being compared, χ^2 will vary widely (and accordingly the value of P by which independence is tested) owing to the size of arbitrary chosen samples. This variation of χ^2 and of P, when there is no natural proportion in the two samples (as sex for example), is often overlooked in interpreting results where χ^2 is really largely determined by the size of the samples compared.

Another important point, to which we may draw attention here, is the relation, often postulated as completely definite, that $\chi^2 = N\phi^2$, where N is the size of the sample. If we couple this relation with the distribution of $\frac{1}{2}\chi^2$ as given by the equation

$$y = y_0' e^{-\frac{1}{2}\chi^2} (\tfrac{1}{2}\chi^2)^{\frac{1}{2}(v-3)} \quad \text{..........} (\alpha),$$

where v is the number of cells, then since the Mean of $\chi^2 = v - 1$ and its variance is $2(v-1)$, we should expect

$$\text{Mean } \phi^2 = \frac{v-1}{N}, \quad \sigma^2_{\phi^2} = \frac{2(v-1)}{N^2} \quad \text{..........} (\beta).$$

Now, if there be κ columns and λ rows in a contingency table, $v = \kappa\lambda$, but the mean value of ϕ^2 and the value of $\sigma^2_{\phi^2}$, even if there be no association, are *not*

$$(\kappa\lambda - 1) N \text{ and } 2(\kappa\lambda - 1)/N \quad \text{..........} (\gamma),$$

see § 17 below of this paper. In the very special case of no association and N large, they only approximate to

$$(\kappa - 1)(\lambda - 1)/N \text{ and } 2(\kappa - 1)(\lambda - 1)/N^2,$$

and even then only agree with the values in (γ), when κ and λ are indefinitely large, but of a definitely lower order than N. A contingency table is hardly likely to be of practical value under such conditions. The fact is that when we are studying the ϕ^2 of a contingency table taken as a sample from an indefinitely large population, successive samples will not have the same marginal totals and the distribution of ϕ^2 is not that of χ^2/N. When we take two series each of definite size, and test their independence by a (χ^2, P) test, we are really dealing with what one of the present writers long ago termed "partial contingency," but it behoves the user to state very precisely what is the origin of the totals of his compared series, and to remember that his P as measuring a degree of independence only applies to repeated comparison of series both of the same totals as the first, and that he cannot generalise as to the degree of dependence which would arise had he used other constant sizes. For the sake of statistical students the senior author of this paper believes it advisable to keep very distinct the usages of χ^2 and ϕ^2, and not obscure a difficult topic by assuming ϕ^2 is merely χ^2/N.

and accordingly we have for the distribution of the difference $m_{\bar{x}_n}' - m_{\bar{x}_n}$ the curve

$$z = \tfrac{1}{2} M T_{Nn(p+1)-\frac{1}{2}} \left(\frac{nNp}{a}(m_{\bar{x}_n}' - m_{\bar{x}_n})\right) \qquad \text{...........(xxxvi)},$$

where M is the total number of cases and p and a refer to the original parent population. We are thus able to test the difference between the means of the means of sets of samples.

Similarly if m_{μ_2}, m_{μ_2}' denote the means of the variances of two sets of samples of n, each N in number, taken from a normal parent distribution, then the distribution of the difference of the variances is given by

$$z = \tfrac{1}{2} M T_{\frac{1}{2}N(n-1)-\frac{1}{2}} \left(\frac{nN}{2} \frac{m_{\mu_2}' - m_{\mu_2}}{M_2}\right) \qquad \text{..............(xxxvii)}.$$

Results (xxxvi) and (xxxvii) may occasionally be useful.

(10) Lastly suppose we have a contingency table, the number of cells being $\kappa \times \lambda$, and a sample of size N be taken from it, then we shall find that the mean square contingency, ϕ^2, of such a sample, *under certain conditions*, obeys the law of distribution

$$y = y_0' (\tfrac{1}{2}\epsilon\phi_1^2)^{p_1} e^{-\frac{1}{2}\epsilon\phi_1^2},$$

but that ϵ is not equal to N_1 nor p_1 to $\frac{1}{2}(v-3)$, where $v = \kappa\lambda$ the number of cells. If two samples having mean square contingencies ϕ^2 and ϕ'^2 with the same number of cells $\kappa \times \lambda$ and of the same size N be drawn under the above-mentioned condition, then the frequency of their difference will be given by

$$y = \tfrac{1}{2} M T_{p_1+\frac{1}{2}} \{\tfrac{1}{2}\epsilon(\phi'^2 - \phi^2)\} \qquad \text{.................(xxxviii)}.$$

The conditions referred to will be discussed in a special section later.

But in many cases ϵ will not equal ϵ', and it is perhaps in practice a more usual problem to determine whether ϕ'^2 may be reasonably supposed to be a sample from the same population as ϕ^2, than to deal with χ'^2 and χ^2. The latter contain the total sizes of the two samples, but the ϕ^2 and ϕ'^2 denoting mean square contingency lead us at once to the problem of whether the coefficients of mean square contingency $\sqrt{\phi'^2/(1+\phi'^2)}$ and $\sqrt{\phi^2/(1+\phi^2)}$ and so the degrees of association in the two samples may be considered as reasonably accordant on the hypothesis of the samples being selected from the same population.

The distribution of $\phi'^2 - \phi^2$ will be considered later. It will be found that, under certain conditions, it obeys the same law of frequency as the distribution of the first product moment coefficient p_{11}.

(11) There is another method of approaching these problems, and only illustrations from known curves or surfaces can tell us which method is generally the more effective or more suitable in a particular type of cases. The whole of our results depend upon quantities $\bar{x}_n, \bar{x}_n'$; μ_2, μ_2'; χ^2, χ'^2; ϕ^2, ϕ'^2, which satisfy a surface which can be thrown into the form

$$w = w_0 e^{-(V+U)} (VU)^p \quad [dV\,dU].$$

So far we have discussed the difference distribution $V - U$ and shown that it is given by a T_m function. We may now discuss the distribution of the ratio $z = V/U$*. Here V and U are independent and can take all values from 0 to ∞. If we integrate out for them and there be M sets of V and U, we find $M = w_0 \Gamma^2(p+1)$ which determines w_0. Now if we consider U and V as measured along two rectangular axes, z = constant gives a line through the origin at slope $\tan^{-1} z$, and if we transfer to z and U as variants, we must integrate keeping z constant from $U = 0$ to ∞ and then for $\tan^{-1} z$ from 0 to $\frac{\pi}{2}$, or z from 0 to ∞.

Thus we find
$$w = w_0 e^{-(1+z)U} U^{2p+1} z^p \quad [dU dz].$$

Put $(1+z) V = \xi$ and keeping z = const., we have
$$w = w_0 e^{\xi} \xi^{2p+1} \frac{z^p}{(1+z)^{2p+2}} \quad [d\xi dz],$$
where ξ goes from 0 to ∞.

Now integrate out for ξ and we find
$$w = w_0 \Gamma(2p+2) \frac{z^p}{(1+z)^{2p+2}} \quad [dz]$$
$$= M \frac{\Gamma(2p+2)}{\Gamma^2(p+1)} \frac{z^p}{(1+z)^{2p+2}} \quad \text{.....................(xxxix).}$$

This is the frequency curve for the distribution of the ratio $z = V/U$ for a population of M ratios. To find its probability integral, we have the measure P_{z_0} that the ratio should be greater than z_0,
$$P_{z_0} = \frac{\Gamma(2p+2)}{\Gamma^2(p+1)} \int_{z_0}^{\infty} \frac{z^p}{(1+z)^{2p+2}} dz.$$

Take $1 + z = \frac{1}{y}$, $dz = -\frac{1}{y^2} dy$, and
$$P_{z_0} = \frac{\Gamma(2p+2)}{\Gamma^2(p+1)} \int_0^{\frac{1}{1+z_0}} (1+y)^p y^p \, dy$$
$$= \frac{B_{\frac{1}{1+z_0}}(p+1, p+1)}{B(p+1, p+1)} = I_{\frac{1}{1+z_0}}(p+1, p+1) \quad \text{..............(xl),}$$

or the incomplete B-function ratio for the value $\frac{1}{1+z_0}$. Here z_0 may be $\bar{x}_n'/\bar{x}_n$, or μ_2'/μ_2, or χ'^2/χ^2, according to the problem with which we are dealing. The quantity $I_x(p', q')$ is that tabled in the *Tables of the Incomplete* B-*function* and from them P_{z_0} can be readily found. In our case we have $p' = q' = p + 1$, or we are confined to the "diagonal" values of that Table.

* Cases of the distribution of V/U have already been considered by R. A. Fisher, V. Romanovsky, and E. S. Pearson with J. Neyman. For the purposes of the present paper we give an independent investigation, which throws the answer back on the *Incomplete B-function Table*. Fisher has provided a table which enables the probability of the ratio U/V to be determined by a transformation of Equation (xxxix).

But the matter would not appear to end here. What we have measured is the improbability that V should exceed $z_0 U$. But we want to measure a certain degree of the probable round $z=1$; we must cut off therefore an equal angle ψ from the OU axis, if ψ be the angle z_0 makes with OV. Clearly $\cot\psi = z_0$ and $\tan\psi = \frac{1}{z_0}$;

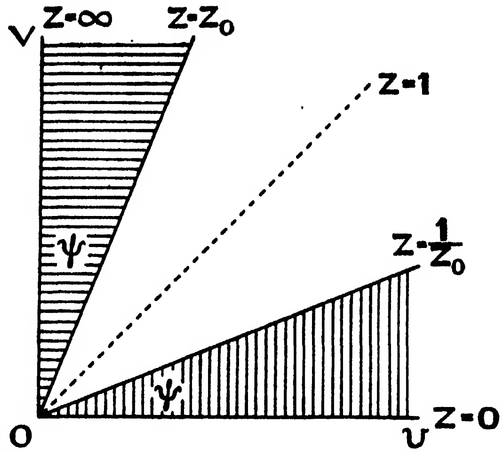


Fig. 2.

therefore the probability that $V/U < \frac{1}{z_0}$ or $U/V < z_0$ is given by

$$P'_{1/z_0} = \frac{\Gamma(2p+2)}{\Gamma^2(p+1)} \int_0^{\frac{1}{z_0}} \frac{z^p}{(1+z)^{2p+2}}\, dz.$$

Taking now $y' = \frac{z}{1+z}$, we have

$$P'_{1/z_0} = \frac{\Gamma(2p+2)}{\Gamma^2(p+1)} \int_0^{\frac{1}{1+z_0}} y'^p (1-y')^p\, dy'$$
$$= Pz_0,$$

as it should, for we have cut off equal areas.

Accordingly the total chance that V/U should exceed z_0 and U/V exceed z_0 is

$$Q_{z_0} = 2I_{\frac{1}{1+z_0}}(p+1, p+1) \quad \text{......................(xli)},$$

which may be taken as a measure of the improbability of the ratios V/U and U/V occurring.

The *Tables of the Incomplete* B-*function* provide Q_{z_0} up to $p=50$, and a small portion of them are reproduced here for comparison. Clearly we need the argument only up to 0·5. (*See* Table II.)

(i) Two means $\bar{x}_n$ and $\bar{x}_n'$ of two samples of size n from a Type VII parent population.

$$Q_{\bar{x}_n'/\bar{x}_n} = 2I_{\frac{1}{1+\bar{x}_n'/\bar{x}_n}}\{n(p+1), n(p+1)\} \quad \text{...............(xlii)}.$$

Thus unless the p of the parental population is known, it has to be approximated to from the samples. Our test would determine whether it was very improbable that the two samples were drawn from the same (a, p) curve.

(ii) Two variances μ_2 and μ_2' of two samples of size n from a supposed normal population.

$$Q_{\mu_2'/\mu_2} = 2I_{\frac{1}{1+\mu_2'/\mu_2}}\{\tfrac{1}{2}(n-1), \tfrac{1}{2}(n-1)\} \quad \text{...............(xliii).}$$

Our test would determine whether it is likely that both samples were taken from normal populations, of the same variance, but not whether those normal populations had the same mean.

(iii) Two χ^2's with the same number of cells v in two samples.

$$Q_{\chi'^2/\chi^2} = 2I_{\frac{1}{1+\chi'^2/\chi^2}}\{\tfrac{1}{2}(v-1), \tfrac{1}{2}(v-1)\} \quad \text{...............(xliv).}$$

(iv) In the case of the mean square contingency we have, under conditions to be discussed in § 16,

$$Q_{\epsilon'\phi'^2/\epsilon\phi^2} = 2I_{\frac{1}{1+\epsilon'\phi'^2/\epsilon\phi^2}}(p_1+1, p_1+1) \quad \text{..................(xlv).}$$

If $\epsilon' = \epsilon$, we have $Q_{\phi'^2/\phi^2}$, but it would be clearly better to be able to provide $Q_{\phi'^2/\phi^2}$ when ϵ' is not equal to ϵ (see p. 304 above), and this will be considered later.

(12) The previous discussion has indicated the necessity of a probability integral table for the $T_m(x)$ curve. We may write

$$S_m(x) = \int_0^x T_m(x)\,dx \quad \text{..........................(xlvi).}$$

A table of $S_m(x)$ has been computed by Miss David (see below). The tabled value being $S_m(x)$, it follows that $\frac{1}{2}(1+\alpha)$ the usual probability integral $= \cdot 5 + S_m(x)$, and $\frac{1}{2}(1-\alpha) = \cdot 5 - S_m(x)$. Hence the probability that an observation will lie outside the limits $\pm x$ is given by $1 - 2S_m(x)$. Considering the case of χ'^2 and χ^2, the difference $\chi'^2 - \chi^2$ and the ratio χ'^2/χ^2 tests will give equal probability when we have

$$1 - 2S_{\frac{1}{2}(v-2)}(\tfrac{1}{2}\chi'^2 - \tfrac{1}{2}\chi^2) = 2I_{\frac{1}{1+\chi'^2/\chi^2}}\{\tfrac{1}{2}(v-1), \tfrac{1}{2}(v-1)\} \quad \text{...(xlvii),}$$

where v is the number of cells under consideration. If we now give v values from 2 upward, and to $\chi'^2/\chi^2 = \lambda$ values from 1 to 100, we are able by means of Table II to find the right-hand side of the equation. Hence by Table I to determine

$$S_{\frac{1}{2}(v-2)}(\tfrac{1}{2}\chi'^2 - \chi^2) = S_{\frac{1}{2}(v-2)}\{\tfrac{1}{2}(\lambda - 1)\chi^2\},$$

and thus χ^2 and $\chi'^2 = \lambda\chi^2$. The curves thus obtained are plotted in Fig. 3. Both the arithmetical work of computing their co-ordinates and the draughtsman's work in producing the diagram were very laborious; the curves asymptote to the axes of χ^2 and χ'^2 being of course symmetrical. They are *not* rectangular hyperbolas, although they might well be described as "hyperboloidea."

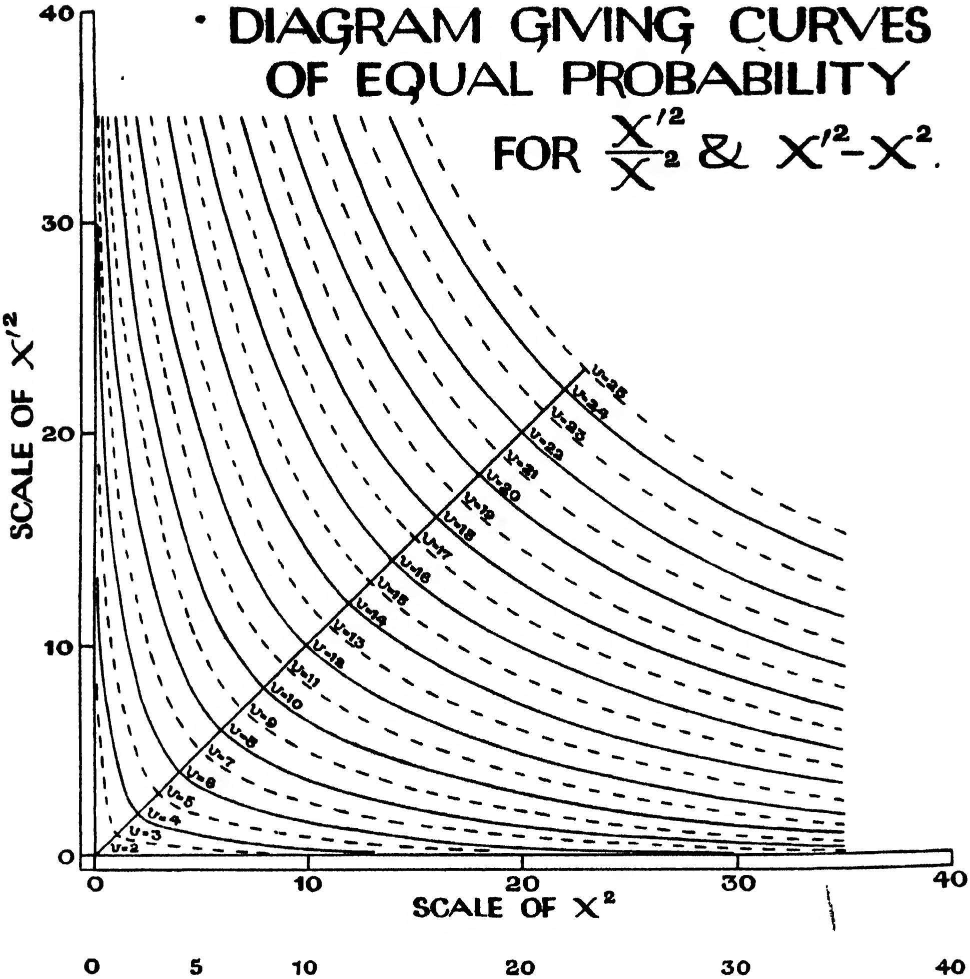


Fig. 3.

When the statistical coefficients by which we enter this diagram are not χ'^2 and χ^2, we must replace them by the following as coordinates:

In case of:

Variances from a normal curve μ_2 and μ_2':

$n\mu_2/\Sigma^2$, $n\mu_2'/\Sigma^2$, where n is the common size of the two samples, and Σ^2 the variance of the parent population or its substitute.

Means from a Type III curve $\bar{x}_n$ and $\bar{x}_n'$:

$\frac{4n}{\sqrt{\beta_1}}\frac{\bar{x}_n}{\Sigma}$, $\frac{4n}{\sqrt{\beta_1}}\frac{\bar{x}_n'}{\Sigma}$, where n is the common size of the two samples; Σ the standard deviation and β_1 belong to the parent population or its substitute. The v in both these cases is $=n$.

Two mean Square Contingencies ϕ_2^2 and $\phi_2'^2$:

$\frac{2p_2}{\rho-1}\phi_2^2$, $\frac{2p_2}{\rho-1}\phi_2'^2$, where ρ is the possible range of $\phi_2^2 = \lambda - 1$, if $\kappa \geqslant \lambda$, $v = 2p_1 + 8$, while p_1 and p_2 are given by Equation (lviii).

Two Correlation Ratios η^2 and η'^2 from a surface of zero correlation:

$(N-n-2)\,\eta^2$, $(N-n-2)\,\eta'^2$, where N is the size of the sample, n the number of arrays, and $v=n$.

Two Correlation Ratios η^2 and η'^2 from a surface of finite correlation:

$2p_2\eta^2$, $2p_2\eta'^2$, where p_1 and p_2 are given by Equation (lxxiv), and $v = 2p_1 + 8$.

Two multiple correlation coefficients R^2 and R'^2:

Replace η^2 and η'^2 by R^2 and R'^2 for the above two cases.

Given a value of v, say $v = 10$, then for values of χ'^2 and χ^2 lying between the curve marked $v = 10$ and its asymptotes the ratio gives a lesser probability than the difference. In other words, the difference test is a more stringent test than the ratio for all points (χ^2, χ'^2) lying *inside* a given v-curve, that is to say a lesser probability of the given hypothesis being correct. Since the area *inside* the curve is always far larger than the area outside the curve—i.e. between the curve and its asymptotes—it would thus appear that the difference test will as a general rule be likely to be the more stringent. But by simply noting the position of the (χ^2, χ'^2) point on the diagram (Fig. 3) it will be found possible to determine which is the more stringent test of a given hypothesis in any particular case.

It may occur to the reader that if the P' or the P corresponding to χ'^2 or χ^2, or indeed both be so small as to render it improbable that either of the compared series have a common origin, it is illogical to test whether χ'^2 and χ^2 have any relation. But a little consideration will show this is not so. For example, let C_1 and C_2 be two processes of inoculation, and let the two processes be applied and the numbers attacked under the two processes be recorded, in each case against a non-inoculated control. Suppose we find in each case from its χ^2 a very slender possibility of the inoculated and control series, being samples of the same parent population, we conclude that inoculation in this matter is of service. But granted this we are at liberty to inquire further whether the two processes of inoculation produce results so divergent that it is unlikely that they themselves could arise from the same population of inoculations. We are really testing whether one or other process is the more effective. Generally our problem will turn on the probability of a difference $\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2$ or a ratio χ'^2/χ^2 greater than the observed occurring. Since no value of χ^2 is "impossible," this probability is a perfectly definite one whatever the actual probabilities of χ'^2 and χ^2 themselves may be. It may be very improbable that the χ'^2 sample belongs to a parent population A, or that the χ^2 sample belongs to a parent population B, neither can be impossible, and accordingly there is no logical reason to hinder us from testing the probability of the combined difference or ratio occurring. All we must be careful about is the interpretation we give to our result.

(13) *Construction of Table I.*

This table was computed in the following manner by Miss F. N. David.

It is known that*

$$\frac{dK_m(x)}{dx} = \frac{m}{x} K_m(x) - K_{m+1}(x) \quad \text{.................(xlviii)},$$

while

$$T_m(x) = \frac{1}{\sqrt{\pi}} \frac{x^m K_m(x)}{2^m \Gamma(m + \frac{1}{2})} \quad \text{........................(xlix)}.$$

Substituting (xlix) in (xlviii) we have, after a slight reduction,

$$T_{m+1}(x) = \frac{2m}{2m+1} T_m(x) - \frac{x}{2m+1} \frac{dT_m(x)}{dx} \quad \text{..................(l)},$$

an equation providing the differential coefficient of $T_m(x)$.

* This follows at once from the equations in *Biometrika*, Vol. XXI. pp. 181 (footnote) and 184.

Integrating from $x=0$ to x, we find

$$S_{m+1}(x)=\frac{2m}{2m+1}S_m(x)-\int_0^x \frac{x}{2m+1}\frac{dT_m(x)}{dx}dx,$$

and integrating the last integral by parts, we conclude that

$$S_{m+1}(x)=S_m(x)-\frac{x}{2m+1}T_m(x) \quad\text{.....................(li).}$$

Now since Dr E. M. Elderton's Table gives $T_m(x)$, we can from a knowledge of $S_m(x)$ find $S_{m+1}(x)$, and thus by repeated use of (li) build up the table of $S_m(x)$. Since we require m to advance by 0·5 intervals, we need to find $S_{\frac{1}{2}}(x)$ and $S_0(x)$.

Now $$S_{\frac{1}{2}}(x)=\int_0^x T_{\frac{1}{2}}(x)\,dx, \quad\text{and}\quad T_{\frac{1}{2}}(x)=\tfrac{1}{2}e^{-x};$$

accordingly

$$S_{\frac{1}{2}}(x)=\tfrac{1}{2}(1-e^{-x}).$$

Thus $S_{\frac{1}{2}}(x)$ could be and was calculated from Glaisher's Table of the Exponential. From this value of $S_{\frac{1}{2}}(x)$ all the values of $S_m(x)$ for $m=1·5, 2·5, \ldots 11·5$ in Table I were computed by (li) in succession.

The value of $S_0(x)=\int_0^x T_0(x)\,dx$ is not so easy to determine, because $T_0(x)$ is infinite when $x=0$, and no quadrature formula is applicable. It was therefore resolved that $S_1(x)=\int_0^x T_1(x)\,dx$ should first be found by quadrature, and then

$$S_0(x)=S_1(x)+xT_0(x)$$

be found from the result, since $L_{x\to 0}\{xT_0(x)\}=0$, which surmounts the difficulty, and thus $S_0(x)$ was determined. The values of $S_1(x)$ obtained by quadratures from $T_1(x)$ were computed by Mr E. C. Fieller, and appear in the column under $m=1·0$ of Table I. The ordinates were taken at intervals of 0·02 from $x=0$ to 0·6, 0·1 of x from $x=0·6$ to 4·0 and after $x=4·0$ up to $x=18·5$ by intervals of 0·5. The work was laborious, the ordinates being calculated to eight figure accuracy, but the areas, given to eight figures, were scarcely to be trusted to the last digit, where there might be an error of 1 to 2. Thus the seventh decimal might sometimes, but rarely, be in error in a unit. For this reason Miss David's Table computed to eight figures was cut down to six for publication. Although the values of $S_m(x)$ for integer values $+0·5$ of m could be obtained with any desired degree of accuracy, those for integer values only depended on a quadrature, which it was difficult to make reliable to eight decimal places. As a matter of fact Miss David's eight figure table was used for all the illustrations which follow, but as linear interpolation was employed* as adequate for the purpose we had in view, we should have got nearly the same final results from the six figure table now published.

Those who have occasion to use the table must be careful to note that from $x=0$ to 4·0, the table advances by 0·1, but from $x=4·0$ to $x=18·0$ by 0·5, and this change must be borne in mind when interpolating into the table.

* In a few cases where the value of x led us to the top of the table higher differences were introduced.

If we are dealing with v categories, $m = \frac{1}{2}(v-2)$, and v is the number of cells indicated by n at the head of the column.

(14) *Construction of Table II.*

The probability of a ratio, e.g. $\chi'^2/\chi^2 = \lambda$, is given by Equation (xlvii), and demands a table of $I_{\frac{1}{1+\lambda}}\{\frac{1}{2}(v-1), \frac{1}{2}(v-1)\}$, where $I_x(p, q)$ is the incomplete B-function ratio, or

$$I_x(p, q) = \frac{\int_0^x x^{p-1}(1-x)^{q-1}\,dx}{\int_0^1 x^{p-1}(1-x)^{q-1}\,dx} \quad \text{.....................(lii).}$$

An important relation between two kinds of B-function ratios may be noted here,

$$I_x(p, p) = \tfrac{1}{2}\{1 + I_{x'}(\tfrac{1}{2}, p)\} \quad \text{.......................(liii),}$$

where $x' = 4(x-\frac{1}{2})^2$.

In the actual table* we should not find $I_{x'}(\frac{1}{2}, p)$ but only $I_{x'}(p, \frac{1}{2})$. The relation between them is

$$I_{x'}(\tfrac{1}{2}, p) = 1 - I_{1-x'}(p, \tfrac{1}{2});$$

thus we modify (liii) and put

$$I_x(p, p) = 1 - \tfrac{1}{2} I_{1-x'}(p, \tfrac{1}{2})\dagger \quad \text{..................(liii } bis\text{),}$$

where $x' = 4(x-\frac{1}{2})^2$.

In actual practice this relationship may be of considerable value as transforming a value of the incomplete B-function from a part of the table where interpolation is difficult to another part where it is easier.

Those who wish to find $I_x(p, p)$ can either use the present paper's Table II or use the values of $P_x(n)$ which have already been published in the Tables of the Probability Integral for Symmetrical Curves issued in *Biometrika*, Vol. XXII. pp. 274—283, or in *Tables for Statisticians* (Part II), pp. 169—178. In this case $P_{x'}(n) = \frac{1}{2}\{1 + I_{x'}(\frac{1}{2}, p)\}$ is actually provided‡ and equals $I_x(p, p)$, where

$$x = \tfrac{1}{2}(1 + \sqrt{x'}).$$

The present Table II renders the discovery of the value of $I_x(p, p)$ very easy. It has not been carried further than $x = {\cdot}50$, for λ only takes values from 1 to ∞.

* Now at press, and shortly to be issued.

† For example, consider $I_{\cdot 7}(6, 6)$; its value taken out directly is ·921,775,209. Now $x = {\cdot}7$, $x' = {\cdot}16$, and $1 - x' = {\cdot}84$, thus the table gives

$$I_{\cdot 84}(6, \tfrac{1}{2}) = {\cdot}156{,}449{,}582.$$

Hence $$1 - \tfrac{1}{2} I_{\cdot 84}(6, \tfrac{1}{2}) = 1 - {\cdot}078{,}224{,}791 = {\cdot}921{,}775{,}209,$$

which is the value of $I_{\cdot 7}(6, 6)$ found directly.

‡ Thus in the example of the previous footnote, we must look out under $\frac{1}{2}(n-1) = 6$, and $x' = {\cdot}16$, and we find ·921,775,2 $= I_{\cdot 7}(6, 6)$ to seven figures.

Table II was extracted from the Incomplete B-function manuscript by Miss M. T. Beer, and, as that table does not go further than 10·5 for the half-unit intervals, the values for $p = 11{\cdot}5$ were computed by Miss Brenda Stoessiger *de novo* in order that the range of Tables I and II might be the same. We have cordially to acknowledge their aid as well as that of Miss M. Kirby for the diagrams, in particular for Fig. 3.

(15) *Illustrations of the Method of using Tables I and II, and of the value of Fig.* 3.

Illustration (i). The following tables are taken from a paper by K. Pearson: *A Study of Trypanosome Strains**.

TABLE i (*a*) OF MEMOIR.

Length of Trypanosomes in Microns.

Goat as Host	12 and under	13	14	15	16	17 and over	Totals
Wild *G. morsitans* Strain	37	55	60	32	12	4	200
Wild Game Strain ...	17	37	73	38	26	9	200

TABLE i (*b*) OF MEMOIR.

Length of Trypanosomes in Microns.

Dog as Host	12 and under	13	14	15	16	17 and over	Totals
Wild *G. morsitans* Strain	17	34	41	40	19	9	160
Wild Game Strain ...	12	31	57	50	24	6	180

If we apply the method of *Biometrika*, Vol. VIII. pp. 250—254, to ascertain whether the Wild *G. morsitans* Strain and Wild Game Strain are probably samples from the same population, we find $\chi'^2 = 17{\cdot}216$ from Table i (*a*) and $\chi^2 = 4{\cdot}745$ from Table i (*b*) leading to $P' = {\cdot}0042$ and $P = {\cdot}4499$ in the two cases respectively for six cells. From the goat as host we should probably argue that the two strains of trypanosomes were different, from the dog as host that they were the same. While the χ'^2 for the goat is very improbable, we must remember that it is not impossible. Two possibilities now arise: (*a*) the two strains are not differentiated by their hosts, (*b*) the two strains are differentiated by their host in the same manner. Are the two χ^2's compatible with each other on either of these hypotheses? We have

$$\tfrac{1}{2}\chi'^2 - \tfrac{1}{2}\chi^2 = 6{\cdot}355 \text{ and } \frac{\chi'^2}{\chi^2} = 3{\cdot}6282.$$

* *Biometrika*, Vol. x. 1914—1915, pp. 117—118.

What do our two tests give us for the probability of compatability in these two χ^2's? We have for the difference test

$$P_{\frac{1}{2}\chi'^2-\frac{1}{2}\chi^2} = 2\{\cdot 5 - S_2(6\cdot 2355)\} = 2\{\cdot 5 - \cdot 493{,}187\}$$
$$= \cdot 0136, \text{ from Table I,}$$
$$Q_{\chi'^2/\chi^2} = 2I_{\cdot 2161}(2\cdot 5, 2\cdot 5) = 2 \times \cdot 0918{,}7173$$
$$= \cdot 1837, \text{ from Table II.}$$

Fig. 3, p. 308, indicates at once that with $\chi'^2 = 17\cdot 216$ and $\chi^2 = 4\cdot 745$, our point is very considerably *inside* the curve for $v = 6$, or without working out the numerical results we know that that difference test will be more stringent than the ratio test. Clearly the ratio test gives us a moderate probability of either (*a*) or (*b*) being the fact, but the difference test suggests that neither hypothesis is correct, or that goat and dog react on the trypanosome strains in different manners. This is in accordance with the P' and P found in the first place for the two tables, but the ratio test being less stringent obscures the first impressions drawn from P' and P. This particular illustration was taken without any knowledge of what the tests would lead to. A similar example, with the χ'^2 smaller, might have made it less easy to draw any definite conclusions from P' and P, while $P_{\frac{1}{2}\chi'^2-\frac{1}{2}\chi^2}$ and $Q_{\chi'^2/\chi^2}$ might one or both give rise to conclusive results.

Illustration (ii). To illustrate the last remark we will take two further tables from Pearson's Memoir on Trypanosomes. They are as follows:

TABLE ii (*a*).

Length of Trypanosomes in Microns.

Goat as Host	11 and under	12	13	14	15	16 and over	Totals
Wild *G. morsitans* Strain	16	21	55	60	32	16	200
Mvera Cattle Strain ...	5	14	22	26	19	14	100

TABLE ii (*b*).

Length of Trypanosomes in Microns.

Dog as Host	11 and under	12	13	14	15	16 and over	Totals
Wild *G. morsitans* Strain	3	14	34	41	40	28	160
Mvera Cattle Strain ...	3	11	27	30	21	8	100

The χ^2 for Table ii (*a*) obtained with a view to testing whether the Wild *G. morsitans* and the Mvera Cattle strains could be samples of the same trypanosome population $= 5\cdot 468$, rendering for six cells a probability $P = \cdot 3646$, or we may say

the Goat as Host cannot be considered as distinguishing between the two strains. We now turn to Table ii (*b*) with the Dog as Host and find $\chi'^2 = 6{\cdot}391$, with the probability $P = {\cdot}2728$. The probability is somewhat less, but far from sufficiently less to enable us to say that the Dog as Host will distinguish between the two strains.

We can now ask on the basis of both tests if it be indifferent whether the difference between the two strains be tested on Goat or Dog?

What is the probability in fact that, in the case of these two strains, the Dog results might have been obtained from the Goat or the Goat results from the Dog?

We have for the difference test

$$P_{\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2} = 2\{{\cdot}5 - S_2({\cdot}4615)\}, \text{ or, from Table I,}$$
$$= 2\{{\cdot}5 - {\cdot}096{,}251\} = {\cdot}8075.$$

Again, for the ratio test, since $\chi'^2/\chi^2 = 1{\cdot}16885$,

$$Q_{\chi'^2/\chi^2} = 2I_{.46107}(2{\cdot}5, 2{\cdot}5), \text{ or, from Table II,}$$
$$= {\cdot}8682.$$

We see that from either test there was high probability of the goat or dog as host being indifferent, but the difference test gives slightly the more stringent result as Fig. 3 *à priori* indicates it must do, although the (χ^2, χ'^2) point is not far removed from the $v = 6$ curve where the two tests give equivalent results.

We might draw from Illustrations (i) and (ii) the conclusion that when the strains are sensibly identical the host is indifferent, but when the strains appear to be different one host may give a more marked reaction than another.

Illustration (iii). Table iii (*a*) below was obtained from the schedules of Pearson's inquiry into the condition of the Polish and Russian Jew immigrants into the East End of London. Table iii (*b*) was adapted from Table VII, p. 255, of Franz Boas's work *Descendants of Immigrants*, New York, Columbia University Press, 1912. The problem to be answered is this: The distributions of Cephalic Indices of the Jewish children born in their adopted country and those born in their land of origin are significantly different. Can this difference be attributed to the same causes in England and in America?

The tables are as follows:

TABLE iii (*a*). (PEARSON'S DATA.)

Cephalic Indices (Central Values).

Male Jewish Boys aged 6 to 15 years	Under 76·95	77·45	78·45	79·45	80·45	81·45	82·45	83·45	84·45	85·45	86·45	87·45	Over 87·94	Totals
Born in England	12	10	13	20	28	31	24	29	20	19	8	7	11	**232**
Born in Eastern Europe ...	3	2	3	2	13	7	7	15	13	11	13	4	12	**105**

TABLE iii (b). (BOAS'S DATA.)

Cephalic Indices (Central Values).

Male Jewish Boys aged 6 to 15 years	Under 77	77·5	78·5	79·5	80·5	81·5	82·5	83·5	84·5	85·5	86·5	87·5	88 and over
Born in America	66	48	121	155	248	263	305	289	244	192	140	69	119
Born in Eastern Europe ...	8	6	10	23	40	47	87	92	93	105	84	82	116

The χ^2 of Table iii (a) is 27·907 corresponding to a P of ·0057, and the χ'^2 of Table iii (b) is 257·399 corresponding to a $P' <$ ·000,0001. Thus the chance of the distribution of the Cephalic Index of Jewish boys born in England being the same as that of Jewish boys born in Eastern Europe is small; the chance that Jewish boys born in America have the same distribution of Cephalic Index as that of Jewish boys born in America is vanishingly small. Boas attributes the difference for America to the influence of the American environment causing the head shape of Jewish children born in America* to approach the Gentile value. Pearson supposes it may be due in the corresponding English case to some admixture of Gentile blood. Whatever the origins of the difference of χ^2's, we may ask how far is there any likelihood of the differences being due to a common cause. In other words, if we took samples of the children of immigrant Jews before and after immigration, what is the chance that two samples will have a difference in their χ^2's equalling or exceeding that observed? The number of cells is 13, and a recourse to Fig. 3 shows us that the point (χ'^2, χ^2) lies well within and away from the $v = 13$ curve; the difference method will therefore be far more stringent than the ratio method.

We have $\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2 = 114{\cdot}746$ and

$$P_{\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2} = 2\,\{{\cdot}5 - S_{5\cdot5}(114{\cdot}746)\}.$$

But Table I shows us that $S_{5\cdot5}(18) = {\cdot}499{,}989$ and $S_{5\cdot5}(114{\cdot}746)$ must be much nearer ·5 than this, or

$$P_{\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2} < {\cdot}000{,}022.$$

Again, $\chi'^2/\chi^2 = 9{\cdot}2234$, and

$$Q_{\chi'^2/\chi^2} = 2I_{10\cdot2234}(6, 6) = 2I_{\cdot0967}(6, 6)$$
$$= {\cdot}000{,}534.$$

Both tests indicate that it is very improbable that the cephalic index divergences have the same cause in America and England, but the difference test is far more stringent.

Thus far we have merely applied the (χ^2, χ'^2) test and drawn an apparent conclusion from it, but in doing so we have really overlooked the warning given in the long footnote to p. 302. While in the case of the trypanosomes in Illustrations (i) and (ii), we have been dealing with total frequencies of much the same order in both

* It is important to note that the mere residence in America is not supposed to modify the head shape of children coming to America. It is the fact of *birth* in America which is credited with the change.

the two sets of tables, so that N' and N of χ'^2 and χ^2 will hardly affect the result; in the present illustration Boas's total number of individuals is nine times Pearson's. Hence, if his proportions remaining the same were reduced to Pearson's total, his χ'^2 would be 28·600 and $\frac{1}{2}(\chi'^2-\chi^2)=0·3465$, giving $S_{5·5}(0·3465)$ instead of $S_{5·5}(114·746)$. We should thus reach

$$P_{\frac{1}{2}\chi'^2-\frac{1}{2}\chi^2}=·9149,$$

or with a high degree of probability conclude that the difference between Pearson's and Boas's Jews and Gentiles could be attributed to a common source.

We do not say that the process here adopted is wholly legitimate, but it does indicate the need for caution in applying the (χ'^2, χ^2) test in either form*, and suggests that a (ϕ'^2, ϕ^2) test may be better applicable when the marginal totals are so different and so arbitrary†.

Illustration (iv). The matter of the preceding illustration may be pursued in a somewhat different direction. The cephalic indices of Jew and Gentile are markedly divergent. If we take as our Gentiles two such closely allied races as English and Swedish, with an almost identical mean index, will either of our tests suffice to indicate a marked difference between the χ^2's for the two series?

Two such series are shown in Tables iv (*a*) and (*b*). Table iv (*a*) is taken from a paper by Nathaniel O. M. Hirsch, entitled: "Cephalic Index of American born Children of three Foreign Groups‡."

Table iv (*b*) is based on Pearson's data for the Jewish Children of East London, and on his data for English School Children.

TABLE iv (*a*). (HIRSCH'S DATA.)

Cephalic Index (Central Values).

Males born in America	Under 74	74·5	75·5	76·5	77·5	78·5	79·5	80·5
Russian Jews	5	2	4	11	21	34	46	65
Swedes ...	17	19	17	19	18	22	12	18

Males born in America	81·5	82·5	83·5	84·5	85·5	86·5	87 and over	Totals
Russian Jews	53	52	49	34	27	17	14	434
Swedes ...	17	10	9	5	4	4	6	197

Here $\chi^2=134·1757$, giving a P for 15 cells $<$ ·000,0005, and there is no practical probability of the two samples coming from the same population.

* $Q_{\chi'^2/\chi^2}=2I._{4939}(6, 6)=·9670$, the difference test being the more stringent.

† The ratio of numbers born in the adopted country to those born in the native land is hardly a "natural" one; for England it is 2·21 and for America 2·85.

‡ *American Journal of Physical Anthropology*, Vol. x. 1910, pp. 79—90, Table I, p. 80.

TABLE iv (*b*). (PEARSON'S DATA.)

Cephalic Index (Central Values).

Males born in England	Under 73·95	74·45	75·45	76·45	77·45	78·45	79·45	80·45
Eastern Jews	—	—	5	7	10	13	20	28
English ...	146	86	167	226	272	342	266	230

Males born in England	81·45	82·45	83·45	84·45	85·45	86·45	86·95 and over	Totals
Eastern Jews	31	24	29	20	19	8	18	232
English ...	188	145	99	61	39	25	21	2313

Here $\chi'^2 = 234{\cdot}6659$ for the 15 cells, and again P' is $< {\cdot}000{,}0005$. Thus both series of data are in accord in indicating that the Jewish male child differs essentially in head shape from either of these series of Gentile male children. But a new problem arises: Is it indifferent whether the Gentiles considered are English or Swedish, what is the probability that χ'^2 and χ^2 could result from two samples drawn from the same Jew-Gentile population?

We will apply first the ratio test. Here $\chi'^2/\chi^2 = 1{\cdot}748{,}954$, and we have, since $v = 15$, and $\frac{1}{2}(v-1) = 7$,

$$Q_{\chi'^2/\chi^2} = 2I_{\cdot 3636}(7, 7) = {\cdot}3077.$$

Thus on the basis of the ratio test, it is not at all improbable that χ'^2 and χ^2 could have arisen from the same population, i.e. it is indifferent whether Swedish or English boys be compared with the Jews.

Now let us consider the difference test. We have $\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2 = 50{\cdot}2451$, and $\frac{1}{2}(v-2) = 6{\cdot}5$. Hence

$$P_{\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2} = 2\{{\cdot}5 - S_{6\cdot5}(50{\cdot}2451)\}.$$

Now $50{\cdot}2451$ is outside the limits of our Table I and $S_{6\cdot5}(18) = {\cdot}499{,}97663$, so $S_{6\cdot5}(50{\cdot}2451)$ has a greater value than this, and we can only say that

$$P_{\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2} < {\cdot}000{,}047.$$

In other words, we should conclude that the difference test strongly points to a divergence between the use of English and Swedes as the Gentile factor, and this would certainly be in accordance with the views of anthropologists, and in particular craniologists. We thus see that the greater stringency of the difference test has led us to a result more in accordance with fact than the ratio test, and

this might apparently justify us in granting it a position at least alongside the ratio test as a statistical method. But here again: Is any conclusion legitimate based upon the series to be compared being as arbitrary in size as the four series of these two tables? These sizes are perfectly arbitrary in the two cases, they were determined by the data each observer chanced to collect, and not by any "natural" proportions. All our analysis tell us is: that if a long series of further experiments were made, always with the same totals for the four series, we should have frequencies determined by the above $P_{\frac{1}{2}\chi'^2-\frac{1}{2}\chi^2}$ and $Q_{\chi'^2/\chi^2}$. But what if we sacrifice the increased accuracy obtained in the case of Hirsch's Jews and Pearson's English and reduced all four series to a common total M? Pearson's formula of 1911* gives

$$\chi^2 = S\,\frac{NN'}{(N+N')^2}\left\{\frac{\left(\frac{f_s}{N}-\frac{f'_s}{N'}\right)^2}{\frac{f+f'}{N+N'}}\right\}.$$

Here the part within curled brackets consists only of proportional frequency and, neglecting influence of random sampling, would remain unchanged if N and N' were modified. Accordingly, if we multiply each χ^2 by

$$\frac{(N+N')^2}{NN'}\times\frac{MM'}{(M+M')^2},$$

we shall reduce it to what would arise if we had the series M and M' instead of N and N'. As there is no reason whatever why we should not take as many Jews as Gentiles, we may put $M=M'$, or the multiplier is $(N+N')^2/4NN'$. For Hirsch's data the multiplier is 1·16424, and for Pearson's 3·01753. Thus we have

$${}_M\chi^2 = 156{\cdot}2127, \quad {}_M\chi'^2 = 708{\cdot}1139,$$

giving $\frac{1}{2}\chi'^2-\frac{1}{2}\chi^2 = 275{\cdot}9506$ and $\chi'^2/\chi^2 = 4{\cdot}53303$. These lead to

$$P_{\frac{1}{2}\chi'^2-\frac{1}{2}\chi^2} = 1-2S_{6\cdot5}(275{\cdot}9506) < {\cdot}000{,}047$$

and much less; and

$$Q_{\chi'^2/\chi^2} = 2I_{\cdot180733}(7,\ 7) = {\cdot}007{,}787.$$

Both probabilities now oppose the suggestion that we are merely comparing Jew and Gentile, they indicate that there is a real difference between Jews compared with Swedes and Jews compared with English. The difference test is, however, much the more stringent.

Illustration (v). We may now turn to another form of application of our method, namely to judicial statistics. We take our data from *Judicial Statistics, England and Wales*, 1925 (Criminal Statistics), Table VII, pp. 68—69, and 1930, Table VII, pp. 56—57; published by H.M. Stationery Office in 1927 and 1932 respectively.

* *Biometrika*, Vol. VIII. p. 252.

Males convicted in England and Wales in Assizes and Quarter Sessions, by Age Groups.

TABLE v (a).

Crimes against the Person.

Year	Under 16	16 and under 21	21 and under 30	30 and under 40	40 and under 60	Over 60	Totals
1925	12	137	369	318	297	45	1178
1930	6	124	326	259	257	36	1008
Totals	18	261	695	577	554	81	2186

TABLE v (b).

Crimes against Property with Violence.

Year	Under 16	16 and under 21	21 and under 30	30 and under 40	40 and under 60	Over 60	Totals
1925	19	591	1059	423	287	47	2426
1930	28	877	1389	506	327	61	3188
Totals	47	1468	2448	929	614	108	5614

Superficially it would appear that Crimes against the Person have decreased at each age, and Crimes against Property with Violence have increased at each age*.

The χ^2 for Table v $(a) = 2{\cdot}0207$ indicating a value of P for six cells of $\cdot 8460$; the samples for the two years might accordingly have arisen from the same population, or we cannot by this test assert a fall in the five years of Crimes against the Person.

Turning to Table v (b) we have $\chi'^2 = 10{\cdot}5304$ with $P' = {\cdot}0626$; this is not absolutely against the 1925 and 1930 results being samples of the same population—if they were, one sample in about 17 would give a greater discrepancy between the two years than the present one—but it does not like the P of the χ^2 of the first table suggest no change in the intensity of crime for the two years.

We may now turn to the usual secondary problem: Is it likely that such changes as are exhibited in the two tables are compatible with a common origin? We ask if the χ'^2 and χ^2 could arise from sampling from a common source. We do not define this common source; it may be that both crimes against the person and against property with violence at each age are decreasing or are stationary or are

* It is to be noted that the data pay no attention to changes in the population of each age group in the five years.

increasing. None of these possibilities is definitely ruled out by an overwhelming improbability, but some of them are not very probable in either one or other case. We have to consider whether χ'^2 and χ^2 are improbable as a result of sampling from a common population.

Our Fig. 3 again shows that the difference test will be the more stringent, but we are not so far from the $v=6$ curve as to believe the two tests will differ much in any inference to be drawn from them. We have at once,

$$P_{\frac{1}{2}\chi'^2-\frac{1}{2}\chi^2} = 2\{\cdot 5 - S_2(4\cdot 25485)\}$$
$$= \cdot 0646, \text{ from Table I.}$$

Again
$$Q_{\chi'^2/\chi^2} = 2I_{\frac{1}{1+5\cdot 21126}}(2\cdot 5,\ 2\cdot 5)$$
$$= 2I_{\cdot 16100}(2\cdot 5,\ 2\cdot 5)$$
$$= \cdot 0942, \text{ from Table II.}$$

The difference test is again more stringent than the ratio test, the former gives odds of 16 to 1 and the latter of 10 to 1 roughly against a common basis for the two tables. On the whole we should probably conclude that the changes noted might not be attributable to a common source, but we should not venture to be dogmatic about such a conclusion. It will be noticed that the $P_{\frac{1}{2}\chi'^2-\chi^2}$ probability is almost the same as the P' for $\frac{1}{2}\chi'^2$, or the difference test pays here little attention to the table in which there is a high probability of the two series being samples of the same population.

Illustration (vi). We will take further data from the same source, and consider whether the changes in the age distributions of those convicted of simple larceny in the years 1925 and 1930 can be contributed to some common cause in the case of the two sexes. The Tables vi (a) and vi (b) are taken from the *Judicial Statistics, England and Wales,* Table X, p. 81, for 1925, and Table X (A), p. 70, for 1930.

Sex and Age of Persons convicted of Simple Larceny in Courts of Summary Jurisdiction (including Juvenile Courts), England and Wales, 1925 *and* 1930.

TABLE vi (a).

Males.

Year	Under 14	14 and under 16	16 and under 21	21 and under 30	30 and under 40	40 and under 60	Over 60	Totals
1925	828	726	2682	4786	2775	2352	347	14496
1930	334	588	2662	4937	3273	2474	390	14658
Totals	1162	1314	5344	9723	6048	4826	737	29154

TABLE vi (*b*).

Females.

Year	Under 14	14 and under 16	16 and under 21	21 and under 30	30 and under 40	40 and under 60	Over 60	Totals
1925	59	54	172	460	489	537	75	1846
1930	17	36	123	455	516	618	61	1826
Totals	76	90	295	915	1005	1155	136	3672

The main feature of the two tables is the decrease in juvenile and the increase in adult thieving. Is the source of this the same for the two series*?

The χ'^2 for Table vi (*a*) is 272·6336, which for $v = 7$ connotes a probability $P' <$ ·000,0001 for the two years being samples of the same population. The χ^2 for Table vi (*b*) is 42·7161, connoting a probability $P <$ ·000,0005 for the two years being samples of the same population. Thus in the case of both males and females there has been a most significant change in the age distributions. We then turn to the problem of whether this change can be attributed to the same source in the two sexes. We have

$$\chi'^2/\chi^2 = 6{\cdot}3825 \quad \text{and} \quad \tfrac{1}{2}\chi'^2 - \tfrac{1}{2}\chi^2 = 114{\cdot}95875.$$

Turning to the ratio test first,

$$Q_{\chi'^2/\chi^2} = 2I_{\cdot 13546}(3, 3)$$
$$= {\cdot}0403.$$

On the ratio test accordingly the odds are about 24 to 1 against two such values, χ'^2 and χ^2, occurring, if there were a common source. We should say therefore that it was unlikely, but not excessively improbable that the age changes in larceny were the same for the two sexes.

We next take the difference test,

$$P_{\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2} = 2\{{\cdot}5 - S_{2\cdot5}(114{\cdot}95875)\}.$$

The value of this $S_{2\cdot5}$ function lies outside our table, but we can say it is considerably greater than $S_{2\cdot5}(18) = {\cdot}499{,}9996$, or we have

$$P_{\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2} < {\cdot}000{,}0008.$$

In other words the difference test shows that the probability of the changes in the two sexes being due to a common source is so vanishingly small that we may safely assert that they are *not* so due.

* The "source" or "sources" may not be changes in the economic or moral state of the population: instead of being of sociological origin the source may lie in police regulations, or in juridical changes; we hazard no suggestion.

Thus the greater stringency of the difference test leads us to a far more definite conclusion than the ratio test does. It may be noted that the marginal totals in Illustrations (v) and (vi) are not the arbitrary sizes of samples, they are the actual populations of criminals caught and convicted and we cannot modify their numbers.

Illustration (vii). We will now apply our methods to a different type of investigation, namely to testing whether the means and variances of small samples are differentiated.

The following data are drawn from Dr M. H. Williams' measurements of boys aged 12 in rural schools in Worcestershire*.

TABLE vii.

Central Heights in Inches of 12 years old Worcestershire Schoolboys.

Group	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
School E	—	1	—	3	2	1	—	4	3	—	1	1	—	—	—	—	—
School F	—	—	—	—	2	3	4	1	4	1	—	—	1	—	—	—	—
Aggregate	2	1	8·5	22·5	24·5	27	45	50	38	32	18·5	4	6	5	—	—	1
Aggregate less E and F	2	—	8·5	19·5	20·5	23	41	45	31	31	17·5	3	5	5	—	—	1
Combined E and F ...	—	1	—	3	4	4	4	5	7	1	1	1	1	—	—	—	—

The means and variances of these groups are as follows:

	Mean	Variance	β_1
School E	54″·0000	7·291,667	—
School F	54″·6875	4·006,510	—
Aggregate	54″·7105	6·539,012	·0085
Aggregate less E and F	54″·7569	6·128,3283	·0384
Combined E and F ...	54″·34375	5·767,2526	·0163

If we have no information as to the β_1 of the aggregate, which sufficiently indicates the symmetry of the distribution, we have (a) the experience that the distribution of stature is very approximately normal for a given age, and (b) the evidence that it is so from the combined samples E and F. The normality of the parent distribution being assumed, we can proceed to test the hypothesis of the equality of the variances for the two schools. The requisite formula is deducible from (xxxi), if we suppose the means of the two samples to be unequal. We have

$$P_{\mu_2' - \mu_2} = 2\left\{\cdot 5 - S_{\frac{1}{2}(n-2)}\left(\frac{n(\mu_2' - \mu_2)}{\mu_2' + \mu_2 + \frac{1}{2}(\bar{x}_n' - \bar{x}_n)^2}\right)\right\}.$$

* "A Statistical Study of Oral Temperatures," by M. H. Williams, M.B., Julia Bell, M.A. and Karl Pearson. *Studies in National Deterioration, Drapers' Company Research Memoirs*, No. IX, Table LXII, p. 109, Cambridge University Press.

Now
$$n=16,\quad \mu_2'-\mu_2=3\cdot285{,}157,$$
$$\mu_2'+\mu_2=11\cdot298{,}177,$$
$$x_n'-x_n=\cdot6875.$$

Hence
$$P_{\mu_2'-\mu_2}=2\{\cdot5-S_7(4\cdot55698)\}$$
$$=\cdot228{,}37 \text{ by Table I.}$$

It is thus quite possible that μ_2' and μ_2 would be equal, if the samples were indefinitely increased.

Let us consider what would happen if we took instead of the variance of the combined samples the variance M_2 of the aggregate of the Worcestershire schools, i.e. 6·539,012. In this case the argument of the S_m function is
$$\frac{n}{2}\frac{\mu_2'-\mu_2}{\mu_2}=\frac{8\times3\cdot285{,}157}{6\cdot539{,}012}=4\cdot019{,}148$$

and
$$P_{\mu_2'-\mu_2}=2\{\cdot5-S_7(4\cdot019{,}148)\}$$
$$=\cdot285{,}09.$$

This makes the equality of the variance in different schools somewhat more probable, but still of much the same order, while if we take the aggregate less Schools E and F, we shall get an intermediate value. This example suggests that it will be adequate in many cases to use the variance of the combined samples in place of the usually unknown variance of the aggregate.

The appropriate equation to use when the samples are of the same size for the ratio is (xliii), or since $\mu_2'/\mu_2=1\cdot819{,}955$, we have
$$Q_{\mu_2'/\mu_2}=2I_{\cdot3546}(7\cdot5,\ 7\cdot5)$$
$$=\cdot258{,}01.$$

The difference test is here slightly more stringent than the ratio test, but neither is incompatible with the hypothesis that the standard deviations may be the same in the two schools. The reader must be cautious in applying Fig. 3 to such a case; he must not use μ_2' and μ_2 as corresponding to χ'^2 and χ^2 and determine the point (μ_2', μ_2) on the diagram. If he did so, he would find that point well outside the $v=16$ curve and so conclude that the ratio test was the more stringent. Equations (xxix) to (xxxi) indicate that the correspondence is between χ'^2 and $n\mu_2'/M_2$ and χ^2 and $n\mu_2/M_2$, or in our particular illustration, the values
$$\frac{16\times7\cdot291{,}667}{5\cdot767{,}2526}\quad\text{and}\quad\frac{16\times4\cdot006{,}510}{5\cdot767{,}2526}\quad\text{or } 20\cdot229 \text{ and } 11\cdot115.$$

Looking this point out on the diagram we find it just inside the $v=16$ curve, showing that the tests will give approximately equal probabilities, but the difference test the smaller one. A like procedure must be adopted in testing on Fig. 3 the relative stringency of other applications of the two methods, i.e. we must inquire from the argument of the S_m function what values correspond to χ'^2 and χ^2. Common factors disappearing in the ratio are apt to mislead, when we apply the difference test.

It will be found generally advisable to use (xxxi) instead of (xxx), unless a preliminary inquiry has settled whether $\bar{x}_n'$ and $\bar{x}_n$ may be considered equal. The tests for this may be practically treated as twofold.

(*a*) If $\bar{x}_1$ and $\bar{x}_2$ be the means of two samples of the same size taken from a normal population of standard deviation Σ, and the samples be perfectly independent, then $\bar{x}_2-\bar{x}_1$ will be distributed normally with standard deviation $(2\Sigma^2/n)^{\frac{1}{2}}$. Accordingly if we form the ratio

$$\zeta=\frac{(\bar{x}_2-\bar{x}_1)\sqrt{n}}{\sqrt{2}\Sigma}$$

we can test for its probability by the integral of the normal curve. When we do not know the parent population, the best value to give to Σ appears to be that of the combined samples. But here a question arises: Should Σ^2 be taken equal to $\frac{1}{2}(\mu_2'+\mu_2)$, which it would be if actually $\bar{x}_2=\bar{x}_1$—our hypothesis—or to

$$\tfrac{1}{2}(\mu_2'+\mu_2)+\tfrac{1}{4}(\bar{x}_2-\bar{x}_1)^2,$$

which is the observed value? We incline to the latter alternative. Accordingly

$$\zeta=\frac{(\bar{x}_2-\bar{x}_1)\sqrt{n}}{\sqrt{\mu_2'+\mu_2+\frac{1}{2}(\bar{x}_2-\bar{x}_1)^2}}=\frac{\cdot 6875\times 4}{\sqrt{11\cdot 534{,}505}}=\cdot 8097$$

in our present case.

The probability of a ratio as large as or larger than this is ·20906, and taking the possibility of either sign for $\bar{x}_2-\bar{x}_1$, we have ·4181 for the chance of a deviation of this size. The test therefore indicates that it would be legitimate to consider the difference between $\bar{x}_2$ and $\bar{x}_1$ as due to random sampling.

If we take for Σ^2 the value 6·539,012 of the aggregate, which would render our theory more accurate*, we find the probability = ·4470, and although this differs to some extent from ·4181, it leads to precisely the same conclusion.

(*b*) We may adopt "Student's *z* test." Here the same assumption of normality for the parent population is made, but we divide $\bar{x}_2-\bar{x}_1$ by the observed standard deviation of the difference $=\sqrt{(\mu_2+\mu_1)}$. Thus in our case

$$z=\frac{\cdot 6875}{\sqrt{11\cdot 298{,}177}}=\cdot 20454,$$

from which we obtain the chance of the mean difference of the samples lying outside $\pm\cdot 6875$ to be ·4406, a value lying between the two values deduced from method (*a*), and thus confirming the result that the schools very probably have the same mean stature for boys of 12.

But to assume this makes no sensible difference in the conclusions already drawn with regard to the variances of the two samples.

Illustration (viii). The following data for pulse rate and oral temperature are taken from the memoir referred to in the last Illustration, Table XI, p. 84.

* Not absolutely so, because the theory assumes that we sample from an indefinitely large population, and that this population is strictly normal.

TABLE viii.

Pulse Rate and Body Temperature in Children.

Pulse Rate

Body Temperature	48–49	50–51	52–53	54–55	56–57	58–59	60–61	62–63	64–65	66–67	68–69	70–71	72–73	74–75	76–77	Totals
A. 98·4°... ...	—	3	7	6	5	1	—	—	—	—	—	—	—	—	—	22
B. 99·4°... ...	—	—	—	2	2	8	5	3	—	2	—	—	—	—	—	22
A and B... ...	—	3	7	8	7	9	5	3	—	2	—	—	—	—	—	44
All Temperatures	3	33	91	189	272	134	88	34	12	7	2	2	—	1	1	869

Before we discuss our samples let us ascertain something about the parent population, of which of course we might have no knowledge, or we might not know that pulse rate curves in children are very skew. We find that the following are the chief constants of the distribution:

$$\text{Mean} = 56{\cdot}8665, \qquad \text{Variance} = M_2 = 10{\cdot}732{,}853,$$

$$\text{Standard Deviation} = \Sigma = 3{\cdot}276{,}103,$$

$$\beta_1 = {\cdot}971{,}479, \qquad \beta_2 = 6{\cdot}159{,}378,$$

the distribution curve is therefore decidedly skew and also markedly leptokurtic.

The distribution of means would have for its constants in the case of samples of 22:

$$\text{Mean} = 56{\cdot}8665, \quad \text{Variance} = \frac{M_2}{n} = {\cdot}487{,}857,$$

$$B_1 = \frac{\beta_1}{n} = {\cdot}044{,}158, \quad B_2 = 3 + \frac{\beta_{2-3}}{n} = 3{\cdot}143{,}608.$$

This is not very widely removed from a normal distribution, and we might well conclude that a normal distribution for the means, as we do not know the true curve, would be sufficient in this case. An examination of the chart of the β_1, β_2 plane* indicates that the corresponding Pearson curve would be Type IV, but that we are so close to the Gaussian point, that the normal curve would be likely to give us quite reasonable results. Hence we are thrown back on (a) or (b) of the last Illustration. We have

$$t = \frac{(\bar{x}_2 - \bar{x}_1)\sqrt{n}}{\sqrt{2}\Sigma} = \frac{5{\cdot}727{,}2727\sqrt{11}}{3{\cdot}276{,}103} = 5{\cdot}798{,}1076 \quad \ldots\ldots\ldots\ldots(\text{liv}).$$

This corresponds to a probability of about ·000,000,007 of such a difference occurring, if we based our investigation on a knowledge of the parent population. It means that samples confined to temperatures 98·4° cannot be considered as randomly chosen or that there is, which we know there to be, a correlation between pulse rate and body temperature. But what are we to do if we do not know the

* See p. 66, *Tables for Statisticians*, Part I.

actual parent population, but suppose it for good reason to be skew? We may put our data as follows:

Pulse-Rate

Group	Mean	Variance	β_1
A. Body Temperature 98·4°...	54·454,546	4·460,055	·024,044
B. Body Temperature 99·4°...	60·181,818	8·997,245	·343,350
A and *B* combined	57·318,182	14·929,063	·237,756
Aggregate of all Temperatures	56·866,513	10·732,853	·971,479

Now assuming that pulse rates can be approximately given by a Type III curve, we turn to equations (xxiii) and (xxiv). Considering first the argument of the T_m function, we see that the first value of Y in (xxiv) would be appropriate if we knew the mean and modal pulse rates, but the latter involves the determination of the mode. Now we can determine these quantities from Equation (viii). We have $p = \frac{4}{\beta_1} - 1 = 4{\cdot}117{,}433$ in the present case, $\bar{x}^* = a\left(1 + \frac{1}{p}\right)$ and $a = \frac{p\sigma}{\sqrt{p+1}}$; hence $\bar{x} = \sigma\sqrt{1+p}$. Thus, for the present case,

$$a = \frac{4{\cdot}117{,}433 \times 3{\cdot}276{,}103}{\sqrt{5{\cdot}117{,}433}} = 5{\cdot}962{,}908.$$

$$\text{Mean} - \text{mode} = \bar{x} - a = \frac{a}{p} = 1{\cdot}448{,}210,$$

and $\tilde{x}_1$ modal pulse rate, accordingly, is given by

$$\tilde{x} = 55{\cdot}418{,}303.$$

Now
$$Y = \frac{n\bar{x}^*(\bar{x}_{n'} - \bar{x}_n)}{\sigma_x^2} = \frac{n\sqrt{1+p}\,(\bar{x}_n' - \bar{x}_n)}{\sigma_x},$$

and for our particular case

$$Y = 87{\cdot}003{,}946.$$

Again, the curve being given by

$$z = \tfrac{1}{2}N\,T_{n(p+1)-\frac{1}{2}}(Y)$$

and $n(p+1) - \frac{1}{2} = 112{\cdot}083{,}526$, we have, for the distribution of Y,

$$z = \tfrac{1}{2}N\,T_{112{\cdot}0835}(Y),$$

and we need the area beyond $87{\cdot}003{,}946$.

But neither $T_m(Y)$ nor $S_m(Y)$ is tabled to such order of the function or to such an ordinate. We turn therefore to p. 297, where we see that after $m = 11{\cdot}5$ the $T_m(Y)$ function coincides with Pearson's Type VII curve. In order to obtain this Type VII curve, we must write, for the $p + \frac{1}{2}$ in (xviii), p. 297, $n(p+1) - \frac{1}{2}$ of the present notation. It then transforms to

$$z = z_0 \frac{1}{\left(1 + \dfrac{Y^2}{4x(p+1)\{n(p+1)+1\}}\right)^{n(p+1)+\frac{1}{2}}} \quad \text{............(lv).}$$

* $\bar{x}$ is measured from start of the pulse rate curve, whereas $\bar{x}_n'$ and $\bar{x}_n$ are absolute values of the means.

In our particular case $n(p+1)+2{\cdot}5=115{\cdot}083{,}526$, but for such values a Type VII curve is for all practical purposes a normal curve, i.e.

$$z=z_0 e^{-\frac{Y^2\{n(p+1)+2\cdot5\}}{4n(p+1)\{n(p+1)+1\}}}.$$

Substituting the value of Y, this becomes

$$z=z_0 e^{-\frac{1}{2}\frac{n(\bar{x}_n'-\bar{x}_n)^2}{2\Sigma^2}\times\frac{n(p+1)+2\cdot5}{n(p+1)+1}},$$

but the latter factor in the exponent is unity to the same degree of accuracy as we have used in passing from the Type VII to the normal curve.

Thus our Bessel function $T_m(Y)$ curve has reduced to the normal curve

$$z=z_0 e^{-\frac{1}{2}\frac{n(\bar{x}_n'-\bar{x}_n)^2}{2\Sigma^2}}=z_0 e^{-\frac{1}{2}\zeta^2} \quad\text{........................(lvi)},$$

precisely the curve from which we obtained our first result that the two means were not compatible with being random samples from the pulse rate population.

It may seem a misfortune that the example chosen does not fall within the range of Table I. We will accordingly try if we have any better luck with the ratio method. This is provided by Equation (xlii). But about this equation an important point must be borne in mind, $\bar{x}_n'$ and $\bar{x}_n$ are not the absolute means, but those means measured from the start of the curve. We must therefore subtract a or $5{\cdot}962{,}908$ from $\tilde{x}$ the modal value or $55{\cdot}418{,}303$, to get the start of the curve which is accordingly at $49{\cdot}455{,}398$ pulse rate*. Thus we have

$$\bar{x}_n'=60{\cdot}181{,}818-49{\cdot}455{,}398=10{\cdot}726{,}420$$

and

$$\bar{x}_n=54{\cdot}454{,}546-49{\cdot}455{,}398=4{\cdot}999{,}148,$$

giving the ratio $\bar{x}_n'/\bar{x}_n=2{\cdot}145{,}650$, or

$$Q_{\bar{x}_n'/\bar{x}_n}=2I_{\cdot31790}(112{\cdot}583{,}526,\ 112{\cdot}583{,}526).$$

Again we find no such high values of the incomplete B-function have been tabled. Writing m for $112{\cdot}583{,}526$, we have

$$Q_{\bar{x}_n'/\bar{x}_n}=2\frac{\int_0^{\cdot31790}x^{m-1}(1-x)^{m-1}dx}{\int_0^1 x^{m-1}(1-x)^{m-1}dx}.$$

Put $x=\frac{1}{2}-x'$ and the transformation gives us

$$Q_{\bar{x}_n'/\bar{x}_n}=2\frac{\int_{\cdot1821}^{\cdot5}(\frac{1}{4}-x'^2)^m dx'}{\int_{-\cdot5}^{+\cdot5}(\frac{1}{4}-x'^2)^m dx'}.$$

* It must be remarked that in this case as in others the ratio as well as the difference test forces us to make appeal by way of knowledge or of hypothesis to the real or supposed parent population. See further on this point p. 328, ftn.

But m is so large that we can safely replace our curve by the normal curve, and thus

$$Q_{\bar{x}_n'/\bar{x}_n} = 2 \frac{\int_{\cdot 1821}^{\cdot 5} e^{-4mx'^2}\, dx'}{\int_{-\cdot 5}^{+\cdot 5} e^{-4mx'^2}\, dx'}.$$

Write $4mx'^2 = \frac{1}{2}\xi^2$, and we have

$$Q_{\bar{x}_n'/\bar{x}_n} = 2 \int_{\sqrt{8m}\times \cdot 1821}^{\sqrt{8m}\times \cdot 5} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\xi^2}\, d\xi.$$

Now $\sqrt{8m} \times \cdot 5 = 14\cdot 93877$ and may be replaced for this integral by ∞ and

$$\sqrt{8m} \times \cdot 1821 = 5\cdot 4407;$$

thus

$$Q_{\bar{x}_n'/\bar{x}_n} = 2 \int_{5\cdot 4407}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\xi^2}\, d\xi \ldots,$$

or, using the probability integral table of the normal curve, $= \cdot 000{,}000{,}05$. Thus we see that on the ratio hypothesis the randomness of the two samples is only slightly less improbable than on the difference test. Both involve a knowledge of the pulse rate distribution. It may be asked, what can we learn without this knowledge? The reply must be that we can only work with the combined A and B as representing to some extent* (!) the general pulse rate distribution. This gives

$$\beta_1 = \cdot 237{,}7557, \quad \sigma_x = 3\cdot 863{,}8146, \quad p + 1 = 16\cdot 823{,}9941$$

and

$$Y = \frac{2n}{\sqrt{\beta_1}} \frac{\bar{x}_n' - \bar{x}_n}{\sigma_x} = 133\cdot 757{,}708,$$

$$m = n(p+1) - \tfrac{1}{2} = 365\cdot 627{,}871,$$

or

$$z = \tfrac{1}{2} N T_{365\cdot 627,871}(133\cdot 757{,}708),$$

a value still farther removed than before from the range of m and Y in the T_m and S_m tables, and still more appropriate to the application of a normal curve. The

* In many cases probably we could appeal to experience as to what sort of value β_1 is likely to take; we know its value for many types of variate, weight, pulse rate, barometer heights, etc. If we do not have any suggestion from experience, the only value we can take is that obtained from the combined samples, even if it appears ludicrous to find β_1 on forty cases. But is it after all *very* much more absurd than finding a variance on twenty cases, which the problem also requires? Again, it may be said that the difference test makes more appeal to or hypotheses about the supposed or known parent population than the ratio test does. But the reader must not forget that the whole theory of χ'^2 in relation to χ^2 is based on the assumption that the relative frequencies of the parent population may be replaced by the relative frequencies of the combined samples. This is clear enough if we approach χ^2 as Pearson has done in *Biometrika*, Vol. VIII. p. 252.

If F_s/M be the relation frequency of the sth category of the parent population M, then the true χ^2 is given by

$$S_s \frac{NN'}{(N+N')^2} \left(\frac{f_s}{N} - \frac{f_s'}{N'}\right)^2 \Big/ \frac{F_s}{M},$$

and it is not till we put $\frac{F_s}{M} = \frac{f_s + f_s'}{N + N'}$, i.e. the relative frequency of the combined samples, that we obtain the value given by various writers for χ^2. In other words, the very basis of the χ^2 method is an appeal to the combined sample relative frequencies as a representation of an infinite parent population. The weakness of this when the samples may consist of 10 to 20 individuals is only too obvious.

normal curve in question will be that given by (lvi). We must now write for ζ in (liv) σ_x for Σ, or 3·863,8146 for 3·276,103, and we obtain

$$\zeta = 4{\cdot}916{,}178,$$

giving a probability

$$P_{\bar{x}_n' - \bar{x}_n} = {\cdot}000{,}000{,}88,$$

not nearly so stringent as the result when we know the parent population, but amply sufficient to indicate that there exists a real difference between the mean pulse rates at temperatures 98·4° and 99·4°.

We have based the above investigation on the assumption that the parent population was a Type III curve with a considerable skewness, but we reach the important conclusion that, even with a β_1 of order 1·0, it will be adequate to apply the normal curve as describing the distribution of samples; with smaller samples and still greater skewness in the parent population, i.e. p smaller, the $m = n(p+1) - 1$ may be small enough to come into the range of our S_m table, or to bring

$$I_x\{n(p+1),\ n(p+1)\}$$

into the range, $n(p+1) = 50$, of the B-function tables.

Unfortunately when we turn to test the standard deviations of the pulse rates of temperatures 98·4° and 99·4° we are somewhat at a loss for a method, for, as far as we are aware, no one has so far found a curve giving the distribution of either the standard deviations or the variances of samples from a Type III curve. All we can provide at present is a curve having the same first four moments as the required curve. To do so would lead us somewhat away from the main topic of the present paper. We can, however, give another illustration of Equation (xxxi) if we assume that we shall not go far wrong by using the Type III distribution of μ_2 to apply approximately to this case.

If we confine our attention to the two samples and their combination, we easily find from the table on p. 326, that

$$Y = \frac{\frac{1}{2} n (\mu_2' - \mu_2)}{\frac{1}{2}(\mu_2' + \mu_2) + \frac{1}{4}(\bar{x}_n' - \bar{x}_n)^2} = \frac{11 \times 4{\cdot}537{,}19004}{14{\cdot}929{,}063} = 3{\cdot}343{,}0825$$

and

$$m = \tfrac{1}{2}(v - 2) = \tfrac{1}{2}(u - 2) = 10.$$

Thus

$$\begin{aligned} P_{\mu_2' - \mu_2} &= 2\{({\cdot}5 - S_{10}(3{\cdot}343{,}0825)\} \\ &= {\cdot}4532 \text{ from Table I.} \end{aligned}$$

If we had used the variance of the total temperature curve, i.e. 10·732,853, we should have found $P_{\mu_2' - \mu_2} = {\cdot}3000$.

Both indicate that it is not unreasonable on our hypothesis to suppose the variance of the two samples the same. Turning now to the ratio test, we have, since $\mu_2'/\mu_2 = 2{\cdot}017{,}2946$,

$$\begin{aligned} Q_{\mu_2'/\mu_2} &= 2I_{\text{[illegible]}}(10{\cdot}5,\ 10{\cdot}5) \\ &= {\cdot}1159 \text{ by Table II.} \end{aligned}$$

This is not entirely opposed to the equality of the two standard deviations, but gives that equality only a fourth of the probability. Thus for the first time in these illustrations the ratio gives the more stringent test. Let us see if this would *à priori* have been indicated by Fig. 3, p. 308; we must take

$$\text{for } \chi'^2\text{: } \frac{n\mu_2'}{M_2} \text{ and for } \chi^2\text{: } \frac{n\mu_2}{M_2},$$

where $M_2 = \frac{1}{2}(\mu_2' + \mu_2) + \frac{1}{4}(\bar{x}_n' - \bar{x}_n)^2 = 14{\cdot}929{,}063$, or $6{\cdot}6293$ and $3{\cdot}2862$ are the required values. The diagram indicates that this point is well outside the curve $v = 22$, and thus the ratio will give the more stringent test.

We have for the purpose of illustration used throughout both tests, but Fig. 3, p. 308, will always enable the investigator to choose *à priori* the test which provides the lesser probability.

(16) We next turn to the important problem of determining from the mean square contingencies whether they may be considered as samples from a common parent population. Now if we have a $\kappa \times \lambda$ contingency table, where $\kappa \geqslant \lambda$, the mean square contingency ϕ_1^2 must lie between 0 and $\lambda - 1$.

Thus as far as a Pearson curve may be considered applicable—this is our first condition—it must theoretically be of the limited range type or of form

$$y = y_0 (\phi_1^2)^{p_1} (\rho - \phi_1^2)^{p_2} \qquad [d\phi_1^2] \text{ ....(lvii)},$$

where $\rho = \lambda - 1$. In the case of such a curve of known range the values of p_1 and p_2 can be determined in terms of the mean ϕ_1^2 and of the variance. Let these be $\bar{\phi}_1^2$ and $\sigma^2_{\phi_1^2}$, then we have

$$\left.\begin{aligned} p_1 + 1 &= \frac{\bar{\phi}_1^2}{\rho}\left(\frac{\bar{\phi}_1^2(\rho - \bar{\phi}_1^2)}{\sigma^2_{\phi_1^2}} - 1\right) \\ p_2 + 1 &= \left(1 - \frac{\bar{\phi}_1^2}{\rho}\right)\left(\frac{\bar{\phi}_1^2(\rho - \bar{\phi}_1^2)}{\sigma^2_{\phi_1^2}} - 1\right) \end{aligned}\right\} \text{ ....(lviii)}.$$

Now when the size of the sample is fairly large, the variance of ϕ_1^2 will be a small quantity compared with the product of the two segments of the range as divided by the mean $\bar{\phi}_1^2$, i.e. $\bar{\phi}_1^2(\rho - \bar{\phi}_1^2)$.

The most unfavourable case for the largeness of either p_1 or p_2 occurs when they are nearly equal and N the size of the sample is small. For example, if $\bar{\phi}_1^2 = {\cdot}5$ and $\sigma^2_{\phi_1^2}$ is of the order ·01 and $\kappa = \lambda = 2$, then $p_1 = p_2 = 11$. (lvii) becomes in this case a Type II curve and the distribution of means of samples from such a curve has not yet been solved in any practically useful manner. But the β_2 for this case is about 2·8 and we should not err greatly by treating the distribution as practically normal*, and then the distribution of the means of $\bar{\phi}_1^2$ would also be normal.

Professor Kondo has made a second experiment for a 3 × 3 table with $N = 250$ (*loc. cit.* in footnote below, pp. 441, 419—420). The number of cells is here again

* Professor Kondo (*Biometrika*, Vol. XXI. pp. 416—418) has dealt with a case of this kind. He has dealt with the *observed* mean and variance of ϕ_1^2 in 804 samples of 100, from a parent population of $\phi_1^2 = {\cdot}5$; here the observed value of $\bar{\phi}_1^2 = {\cdot}499{,}8005$ rendered the curve slightly skew.

very limited, but any one who has endeavoured to take several hundred samples of contingency for 2×2 tables will appreciate the amount of labour involved. In this second experiment p_2 is large, and it would be adequate practically to replace (lvii) by

$$y = y_0 (\phi_1^2)^{p_1} e^{-\frac{p_2}{\rho}\phi_1^2} \quad \text{.......................(lix).}$$

If p_1 were the larger, we should have to measure our variate ϕ_1^2 from the other end of the range and it would be the term with power p_1 which would be replaced by the exponential. If therefore either p_1 or p_2 be large, we can write our distribution of ϕ_1^2

$$y = y_0' (\tfrac{1}{2}\epsilon\phi_1^2)^{p_1} e^{-\frac{1}{2}\epsilon\phi_1^2} \quad \text{.......................(lix } \textit{bis}\text{),}$$

where $\frac{1}{2}\epsilon = \frac{p_2}{\rho}$, and p_1 and p_2 are given by (lviii). In this case the distribution of $\phi_1'^2 - \phi_1^2$, if the samples are of the same size and the same number of cells, is given by

$$z = \tfrac{1}{2} M T_{p_1 + \frac{1}{2}} \{\tfrac{1}{2}\epsilon(\phi_1'^2 - \phi_1^2)\} \quad \text{.......................(lx),}$$

and the probability of a difference as great or greater is given by

$$P_{\phi_1'^2/\phi_1^2} = 1 - 2S_{p_1+\frac{1}{2}} \{\tfrac{1}{2}\epsilon(\phi_1'^2 - \phi_1^2)\} \quad \text{..................(lxi).}$$

The corresponding probability of the ratio $\phi_1'^2/\phi_1^2$ is given by

$$Q_{\phi_1'^2/\phi_1^2} = 2I_{\frac{1}{1+\phi_1'^2/\phi_1^2}}(p_1+1, p_1+1) \quad \text{..................(lxii).}$$

Thus the difference test involves the determination of one more constant p_2 than the ratio test with p_1 only. But this does not much increase the labour, as p_1 already involves a knowledge of both $\bar{\phi}_1^2$ and $\sigma^2_{\phi_1^2}$, which are the laborious quantities to determine. The values of these quantities to a *second* approximation are given in the memoir by Professor Kondo, already referred to. We will now discuss special cases.

(17) The problem to be discussed in this section is to find the probability that two values $\phi_1'^2$ and ϕ_1^2 of mean square contingency could have been obtained for samples of the same size, N, the same number of cells, $\kappa \times \lambda$, with a parent population of zero contingency.

In this case we know that approximately*

$$\bar{\phi}_1^2 = (\kappa-1)(\lambda-1)\frac{1}{N}\left(1 + \frac{1}{N}\right) \quad \text{..............................(lxiii),}$$

$$\sigma^2_{\phi_1^2} = \frac{2}{N^2}(\kappa-1)(\lambda-1)\left\{1 - \frac{(\kappa-1)(\lambda-1)}{N}\left(1 + \frac{1}{2N}\right)\right\} \quad \text{...(lxiv).}$$

We might compute these for any given case N, κ, λ, and then substitute in (lviii). But as the approximation is only of the second order, we can save workers some trouble by making the substitution, algebraically neglecting in the process terms which we are not warranted in retaining.

* Kondo, *loc. cit.* p. 408.

Remembering that $\rho = \lambda - 1$, if $\lambda \leqslant \kappa$, and calling F the factor common to p_1 and p_2, we have

$$F = \frac{\bar{\phi}_1^2(\rho - \bar{\phi}_1^2)}{\sigma^2_{\phi_1^2}} - 1 = \tfrac{1}{2}N(\lambda - 1)\left\{1 + \frac{\nu}{N} + \frac{(\kappa-1)(\lambda-1)}{N^2}(\nu + \tfrac{1}{2})\right\} \quad \ldots\text{(lxv)},$$

where $\nu = (\kappa - 1)(\lambda - 1) - (\kappa - 2) - \dfrac{2}{\lambda - 1}$.

Now

$$p_1 + 1 = \frac{\bar{\phi}_1}{\rho} \times F = \tfrac{1}{2}(\kappa - 1)(\lambda - 1)\left(1 + \frac{1}{N}\right)\left\{1 + \frac{\nu}{N} + \frac{(\kappa-1)(\lambda-1)}{N^2}(\nu + \tfrac{1}{2})\right\}$$

$$= \tfrac{1}{2}(\kappa - 1)(\lambda - 1)\left\{1 + \frac{\nu + 1}{N} + \frac{(\kappa-1)(\lambda-1)(\nu + \frac{1}{2}) + \nu}{N^2}\right\} \quad \ldots\text{(lxvi)}.$$

For the extreme case of $N \rightarrow \infty$, we see that $p_1 + 1 \rightarrow \frac{1}{2}(\kappa - 1)(\lambda - 1)$, or we have for the ratio test

$$Q_{\phi_1'^2/\phi_1^2} = 2I_{\frac{1}{1+\phi_1'^2/\phi_1^2}}\{\tfrac{1}{2}(\kappa - 1)(\lambda - 1), \tfrac{1}{2}(\kappa - 1)(\lambda - 1)\}.$$

For the χ^2 test we found

$$Q_{\chi'^2/\chi^2} = 2I_{\frac{1}{1+\chi'^2/\chi^2}}\{\tfrac{1}{2}(v - 1), \tfrac{1}{2}(v - 1)\},$$

so that we cannot pass to the ϕ_1^2 test by writing merely $v = \kappa\lambda$. This would only be true if κ and λ were very large indeed, which would be very rare in practical work. For the like reason $p_1 + \frac{1}{2}$ can only be taken as $\frac{1}{2}(v - 2)$, where $v = \kappa\lambda$, when not only N is very large but $(\kappa + \lambda - 2)/(\kappa\lambda)$ is a very small fraction. For example, if $\kappa = \lambda = 10$, or a contingency table of 100 cells, which is very unusual, we can hardly assert that 18/100 is a very small fraction.

Next turning to p_2 we have

$$p_2 + 1 = \tfrac{1}{2}N(\lambda - 1)\left\{1 - \frac{\kappa - 1}{N}\left(1 + \frac{1}{N}\right)\right\}\left(1 + \frac{\nu}{N} + \frac{(\kappa-1)(\lambda-1)}{N^2}(\nu + \tfrac{1}{2})\right)$$

$$= \tfrac{1}{2}N(\lambda - 1)\left(1 + \frac{\nu - (\kappa - 1)}{N} + \frac{(\kappa - 1)\{(\lambda - 2)(\nu + \frac{1}{2}) - \frac{1}{2}\}}{N^2}\right) \quad \ldots\ldots\ldots\text{(lxvii)}.$$

Illustration (ix). Suppose $\kappa = \lambda = 5$ and $N = 400$, what is the probability that $\phi_1^2 = {\cdot}02$ and $\phi_1'^2 = {\cdot}07$ could both be drawn as samples of N from a population of zero contingency?

Here we find from (lxiii) and (lxiv) that $\bar{\phi}_1^2 = {\cdot}040{,}100$, and $\sigma^2_{\phi_1^2} = {\cdot}0001{,}9199$, while $\rho = \lambda - 1 = 4$.

We may now proceed to find p_1 and p_2 either from (lviii), or from (lxvi) and (lxvii). We cannot say that (lviii) is more accurate than (lxvi) and (lxvii), because in (lviii) we retain terms which are of an order we have neglected in (lxiii) and (lxiv), and which ought to be neglected because they have been neglected in passing from (lvii) to (lix). Working with (lviii) we find first from (lxiii) and (lxiv),

$$\bar{\phi}_1^2 = {\cdot}040{,}100, \quad \sigma^2_{\phi_1^2} = {\cdot}0001{,}9199,$$

whence the common factor $F = 826{\cdot}0847$. Further $\bar{\phi}_1^2/\rho = {\cdot}010{,}025$, and

$$1 - \frac{\bar{\phi}_1^2}{\rho} = {\cdot}989{,}975,$$

leading by (lviii) to

$$p_1 = 7{\cdot}2815 \quad \text{and} \quad p_2 = 816{\cdot}7803.$$

Working with (lxii) we have at once

$$\nu = 16 - 3 - \tfrac{1}{2} = 12{\cdot}5,$$

whence $$F = 800\,(1 + {\cdot}03125 + {\cdot}0013) = 826{\cdot}0400.$$

Thus $$p_1 = 7{\cdot}2811 \quad \text{and} \quad p_2 = 816{\cdot}7589.$$

We see therefore that the two methods accord well, and further that p_2 is so large that (lix *bis*) wil' adequately express the distribution, where $\frac{1}{2}\epsilon = 204{\cdot}1897$.

Finally, by (lxi), the probability of the occurrence of the difference

$$\phi_1'^2 - \phi_1^2 = {\cdot}05$$

is given by $$P_{\phi_1'^2 - \phi_1^2} = 1 - 2S_{7\cdot7815}(10{\cdot}2095).$$

And again $$Q_{\phi_1'^2/\phi_1^2} = 2I_{\cdot222,222}(8{\cdot}2815,\, 8{\cdot}2815).$$

Both involve a twofold interpolation with regard first to the arguments, and secondly to the orders of the functions. For the purpose of appreciating the probabilities the hyperbolic formula* will suffice. From Table I we find

$$S_{7\cdot7815}(10{\cdot}2095) = {\cdot}4920{,}0818$$

and thus $$P_{\phi_1'^2 - \phi_1^2} = {\cdot}01598.$$

Again from Table II we have

$$I_{\cdot222,222}(8{\cdot}2815,\, 8{\cdot}2815) = {\cdot}008{,}324{,}230,$$

and accordingly $$Q_{\phi_1'^2/\phi_1^2} = {\cdot}01665.$$

Such values would only lead us to say that it is not very probable that $\phi_1'^2$ and ϕ_1^2 were samples from a population having zero contingency. Here, as so often, the difference test is, if only slightly, still more stringent than the ratio test.

In order to determine this *à priori* from our diagram, Fig. 3, p. 308, we must first find v from the relation $\frac{1}{2}(v-2) = 7{\cdot}7815$, or $v = 17{\cdot}563$. The curve corresponding to this is nearly mid-way between the curves for $v = 17$ and 18. Corresponding to χ^2 and χ'^2 we have $\epsilon\phi_1^2$ and $\epsilon\phi_1'^2$, or 8·1676 and 28·5866. This point is just inside the $v = 18$ curve and so clearly within the $v = 17{\cdot}563$ curve; thus the difference test is the more stringent.

(18) The previous section indicates that the solution of the problem of two mean square contingencies arising as samples from the same population is relatively easy, if that population has zero contingency. This follows from the fact that the approximate expressions for $\bar{\phi}_1^2$ and $\sigma_{\phi_1^2}$ are known and relatively simple. In the case where the contingency is not zero the problem is much harder, as although approximate expressions are known for $\bar{\phi}_1^2$ and $\sigma^2_{\phi_1^2}$ they are laborious to determine in particular cases.

* *Tables for Statisticians and Biometricians*, Part II. Formula (a), p. xviii, i.e.

$$z_{\theta\chi} = \phi\psi z_{00} + \phi\chi z_{01} + \theta\psi z_{10} + \theta\chi z_{11} \quad \text{......(lxviii)}.$$

Illustration (x). We will first illustrate the method in a particular example provided by Kondo and then consider what is needed in order that the values of $\bar{\phi}_1^2$ and $\sigma^2_{\phi_1^2}$ might be obtained more readily.

Kondo took 250 samples of size 200 from an infinite population of a 3 × 3 table with the following proportional frequencies*:

·0831 ·1032 ·0335	·0786 ·2864 ·1235	·0270 ·0862 ·1785	·1887 ·4758 ·3355
·2198	·4885	·2917	1·0000

From this table Kondo computes

$$\bar{\phi}_1^2 = \cdot 206{,}503, \quad \sigma^2_{\phi_1^2} = \cdot 0043{,}3796.$$

Clearly $\rho = \lambda - 1 = 2$, and thus

$$\bar{\phi}_1^2/\rho = \cdot 103{,}2515 \quad \text{and} \quad \rho - \bar{\phi}_1^2 = 1\cdot 793497.$$

Using Equations (lviii) we find

$$p_1 = 7\cdot 712{,}064, \quad p_2 = 74\cdot 665{,}056,$$

only differing slightly from Kondo's values†. p_2 is considerable, and we take it that the distribution curve of ϕ_1^2 may be reasonably represented by (lix *bis*), or

$$y = y_0' \tfrac{1}{2} (74\cdot 665{,}056\, \phi_1^2)^{7\cdot 712{,}064} e^{-\frac{1}{2}(74\cdot 665{,}056)\phi_1^2}.$$

Accordingly the chance of a difference as great as or greater than $\phi_1'^2 - \phi_1^2$ is

$$P_{\phi_1'^2 - \phi_1^2} = 1 - 2S_{8\cdot 212{,}064}\{\tfrac{1}{2} 74\cdot 665{,}056\,(\phi_1'^2 - \phi_1^2)\},$$

and the chance of a ratio at least as great as $\phi_1'^2/\phi_1^2$ is

$$Q_{\phi_1'^2/\phi_1^2} = 2I_{\frac{1}{1+\phi_1'^2/\phi_1^2}}(8\cdot 712{,}064,\ 8\cdot 712{,}064).$$

Now Kondo has given the 250 values of ϕ_1^2 which he obtained for his samples; the largest of these is ·370,502 and the least ·054,201‡. Let us inquire what is the chance of such a difference occurring in samples of 200 taken from the table above.

The difference $\phi_1'^2 - \phi_1^2 = \cdot 316{,}301$ and the ratio $\phi_1'^2/\phi_1^2 = 6\cdot 835{,}704$; hence we require

$$S_{8\cdot 212{,}064}(11\cdot 808{,}316),$$

and

$$I_{\cdot 127{,}631}(8\cdot 712{,}064,\ 8\cdot 712{,}064).$$

Using the hyperbolic interpolation formula (lxviii), we find

$$S_{8\cdot 212{,}064}(11\cdot 808{,}316) = \cdot 496{,}3113,$$

and

$$I_{\cdot 127{,}631}(8\cdot 712{,}064,\ 8\cdot 712{,}064) = \cdot 000{,}111{,}08.$$

Accordingly $\quad P_{\phi_1'^2 - \phi_1^2} = \cdot 007{,}3774 \quad \text{and} \quad Q_{\phi_1'^2/\phi_1^2} = \cdot 000{,}222{,}16.$

* *Biometrika*, Vol. xxi. p. 411. † *Biometrika*, Vol. xxi. p. 419. ‡ *Ibid.* p. 412.

Referring to Fig. 3, p. 308, we have to enter with $\epsilon\phi_1'^2 = 27\cdot66$, $\epsilon\phi_1^2 = 4\cdot05$ and $v = 18\cdot424$, or since this point lies well below even the $v = 18$ curve the ratio test is here the more stringent. With 250 samples of 200 we should expect by the difference test 1·8 cases with a ϕ_1^2 difference as great as or greater than ·316. Kondo experimentally found one, but a second ·313 runs it close. By the ratio test we should expect only ·0555, say ·06 of a case in 250 samples. Thus while the ratio test is the more stringent in this particular case, the difference test accords better with Kondo's experience, and would determine more accurately the range of ϕ_1^2 in such an experiment. It may be remarked that neither Kondo's maximum nor his minimum ϕ_1^2's are outlying values; they run:

At top ·370,502, ·367,532, ·362,669, ·343,805,

and at bottom ·054,201, ·063,055, ·080,840, ·098,408.

Thus the two methods could not be brought more into accordance by the omission of an exceptional outlyer. Assuming the arithmetic to be correct, and it has been carefully checked, this case seems to be of importance as an indication of the value of the difference test when subjected to experimental verification.

The above illustration shows that there is no difficulty in applying the difference test to two values of ϕ_1^2, if the parent population with a definite contingency be supposed known and the two samples have the same size and the like cell distribution, *provided* $\bar{\phi}_1^2$ and $\sigma^2_{\phi_1^2}$ can be determined. When we have no real or hypothetical parent population, our only method is to suppose the parent population to be obtained from the combined samples which are used to give the relative frequencies.

(19) The problem remains as to what can be done to simplify the labour of obtaining $\bar{\phi}_1^2$ and $\sigma^2_{\phi_1^2}$ in the case of contingency tables with a reasonably large size of sample. Now Professor Kondo has shown* that, to a second approximation,

$$\left.\begin{aligned}\bar{\phi}_1^2 &= \tilde{\phi}_1^2 + \frac{\psi_1}{N} + \frac{\psi_2}{N^2}\\ \sigma^2_{\phi_1^2} &= \frac{f_1}{N} + \frac{f_2}{N^2} - \left(\frac{\psi_1}{N} + \frac{\psi_2}{N^2}\right)^2\end{aligned}\right\} \quad \text{.....................(lxix)},$$

where ψ_1, ψ_2, f_1, f_2 are functions of the *relative* frequencies of the parent population —i.e. of its cell frequencies on the basis of a total frequency of unity†—and $\tilde{\phi}_1^2$ is the mean squared contingency of the same population‡.

Now if we substitute these values in (lviii), the expressions for p_1 and p_2 become very complicated, even if we only retain the first two terms in the expansion. We get simple results if we retain only the leading term. In this case

$$\left.\begin{aligned}p_1 &= \frac{N(\tilde{\phi}_1^2)^2(\rho - \tilde{\phi}_1^2)}{\rho f_1} - 1\\ p_2 &= \frac{N\tilde{\phi}_1^2(\rho - \tilde{\phi}_1^2)^2}{\rho f_1} - 1\end{aligned}\right\} \quad \text{.......................(lxx)}.$$

* *Biometrika*, Vol. XXI. p. 418.

† They are in fact the chances that an individual will be drawn from each particular cell.

‡ $\bar{\phi}_1^2$ is of course the mean of the mean squared contingency of samples and only approaches $\tilde{\phi}_1^2$ as N is increased.

These expressions do not contain ψ_1, ψ_2 or f_2, which would not therefore require calculation. How far can (lxx) be used instead of (lviii) and the fuller but still incomplete results (lxix)? This cannot be determined till far more experimental work has been undertaken on the two sets of approximations. We may place here the formulae from which $\tilde{\phi}_1^2$ and f_1 are to be determined. Let c_{pq} be the chance that an individual will be drawn from the pth row qth column, $c_{p.}$ the chance that it may be drawn from anywhere in the pth row and $c_{.q}$ from anywhere in the qth column. Let $C_{pq} = c_{pq}/(c_{p.}\, c_{.q})$ and $C_{p.} = \overset{q=\lambda}{\underset{q=1}{S}} C_{pq}$, i.e. the sum of the C_{pq} for the pth row; in the same manner let $C_{.q} = \overset{p=\kappa}{\underset{p=1}{S}} C_{pq}$ or be the sum of C_{pq} for the qth column. Then

$$\tilde{\phi}_1^2 = \Sigma(C_{pq}) - 1 \quad \text{.............................(lxxi)},$$

where Σ denotes a summation for the whole tables, and

$$f_1 = 4\Sigma\left(\frac{C^2_{pq}}{c_{pq}}\right) - 3\overset{p=\kappa}{\underset{p=1}{S}}\left(\frac{C^2_{p.}}{c_{p.}}\right) - 3\overset{q=\lambda}{\underset{q=1}{S}}\left(\frac{C^2_{.q}}{c_{.q}}\right) + 2\Sigma\frac{C_{pq}C_{p.}C_{.q}}{c_{pq}} \text{......(lxxii)}.$$

A sample of the actual working required to obtain an f_1 is provided on the following page, it is for the table of Kondo's on p. 334 of this paper. It is considerable, but not prohibitive for a table of 3×3 cells. We can now illustrate the approximate formulae (lxx) on this case where

$$\tilde{\phi}_1^2 = \cdot 188{,}8925, \quad \rho = 2, \quad f_1 = \cdot 804{,}0440.$$

These give us $p_1 = 7\cdot 0370$ and $p_2 = 76\cdot 0590$, whereas when we use (lviii) with the fuller values given by Kondo for $\bar{\phi}_1^2$ and $\sigma^2_{\phi_1^2}$ we have $p_1 = 7\cdot 7121$ and $p_2 = 74\cdot 6651$ *.

We shall have $\frac{1}{2}\epsilon = 38\cdot 0295$, and accordingly

$$P_{\phi_1'^2 - \phi_1^2} = 1 - 2S_{7\cdot 5870}(12\cdot 0288).$$

From Table I we find

$$S_{7\cdot 5870}(12\cdot 0288) = \cdot 4971{,}7440$$

and accordingly $P_{\phi_1'^2 - \phi_1^2} = \cdot 00565$.

Thus in 250 samples we might expect 1·4 occasions on which a difference as great as or greater than that observed would arise. Accordingly we have actually approached nearer to the experimental result by using the less approximate values of $\bar{\phi}_1^2$, and $\sigma^2_{\phi_1^2}$. For such a table as we are dealing with, it is clear that Equations (lxx) will amply suffice for practical purposes.

(20) We can still further extend the usefulness of our $T_m(Y)$ function and its probability integral to many cases where we wish to investigate the difference between two squared correlation ratios or two squared multiple correlation coefficients. If these be η^2 and R^2, those quantities can only vary between 0 and 1, and an appropriate curve for their distributions will be

$$\left.\begin{aligned} y &= y_0(\eta^2)^{p_1}(1-\eta^2)^{p_2} \\ \text{or} \quad y &= y_0(R^2)^{p_1}(1-R^2)^{p_2} \end{aligned}\right\} \text{......................(lxxiii)}.$$

* Kondo gives $f_1 = \cdot 801{,}842$, but we have failed to find a slip in our arithmetic. This leads to $p_1 = 7\cdot 059$, $p_2 = 76\cdot 271$ and $P_{\phi_1'^2 - \phi_1^2} = \cdot 00490$.

Actual Working of an f_1.

	$c_{11}=\cdot 0831$	$c_{12}=\cdot 0786$	$c_{13}=\cdot 0270$	$c_{1\cdot}=\cdot 1887$
c^2_{1q}	·0069,0561	·0061,7796	·0007,2900	$C_{1\cdot}=S(C_{1q})=\cdot 246{,}7601$
$c_{1\cdot}\,c_{\cdot q}$	·0414,7626	·0921,7995	·0550,4379	$C^2_{1\cdot}=\cdot 0608{,}9055$
C_{1q}	·166,4955	·067,0206	·013,2440	$\dfrac{C^2_{1\cdot}}{c_{1\cdot}}=\cdot 322{,}6844$
C^2_{1q}	·0277,2075	·0044,9176	·0001,7540	
C^2_{1q}/c_{1q}	·333,5830	057,1471	·006,4963	$S_{1\cdot}\left(\dfrac{C^2_{1q}}{c_{1q}}\right)=\cdot 397{,}2264$
C_{1q}/c_{1q}	2·003,5560	·852,6794	·490,5185	
	$c_{21}=\cdot 1032$	$c_{22}=\cdot 2864$	$c_{23}=\cdot 0862$	$c_{2\cdot}=\cdot 4758$
c^2_{2q}	·0106,5024	·0820,2496	·0074,3044	$C_{2\cdot}=S(C_{2q})=\cdot 508{,}2788$
$c_{2\cdot}\,c_{\cdot q}$	·1045,8084	·2324,2830	·1387,9086	$C^2_{2\cdot}=\cdot 2583{,}4734$
C_{2q}	·101,8374	·352,9044	·053,5370	$\dfrac{C^2_{2\cdot}}{c_{2\cdot}}=\cdot 542{,}9747$
C^2_{2q}	·0103,7086	·1245,4152	·0028,6621	
C^2_{2q}/c_{2q}	·100,4928	·434,8517	·033,2507	$S_{2\cdot}\left(\dfrac{C^2_{2q}}{c_{2q}}\right)=\cdot 568{,}5952$
C_{2q}/c_{2q}	·986,7965	1·232,2081	·621,0789	
	$c_{31}=\cdot 0335$	$c_{32}=\cdot 1235$	$c_{33}=\cdot 1785$	$c_{3\cdot}=\cdot 3355$
c^2_{3q}	·0011,2225	·0152,5225	·0318,6225	$C_{3\cdot}=S(C_{3q})=\cdot 433{,}8536$
$c_{3\cdot}\,c_{\cdot q}$	·0737,4290	·1638,9175	·0978,6535	$C^2_{3\cdot}=\cdot 1882{,}2890$
C_{3q}	·015,2184	·093,0629	·325,5723	$\dfrac{C^2_{3\cdot}}{c_{3\cdot}}=\cdot 561{,}0401$
C^2_{3q}	·0002,3160	·0086,6070	·1059,9732	
C^2_{3q}/c_{3q}	·006,9134	·070,1271	·593,8226	$S_{3\cdot}\left(\dfrac{C^2_{3q}}{c_{3q}}\right)=\cdot 670{,}8631$
C_{3q}/c_{3q}	·454,2806	·753,5457	1·823,9345	
$c_{\cdot q}$	$c_{\cdot 1}=\cdot 2198$	$c_{\cdot 2}=\cdot 4885$	$c_{\cdot 3}=\cdot 2917$	$c_{\cdot\cdot}=1\cdot 0000$
$C_{\cdot q}$	·283,5513	·512,9879	·392,3533	$\Sigma(C_{pq})=1\cdot 188{,}8925$; thus $\phi^2=\cdot 188{,}8925$
$C^2_{\cdot q}$	·0804,0134	·2631,5659	·1539,4111	$S_p\left(\dfrac{C^2_{p\cdot}}{c_{p\cdot}}\right)=1\cdot 426{,}6992$; $S_q\left(\dfrac{C^2_{\cdot q}}{c_{\cdot q}}\right)=1\cdot 432{,}2344$
$\dfrac{C^2_{\cdot q}}{c_{\cdot q}}$	·365,7932	·538,7034	·527,7378	$\Sigma\left(\dfrac{C^2_{pq}}{c_{pq}}\right)=S_{1\cdot}+S_{2\cdot}+S_{3\cdot}=1\cdot 636{,}6847$
$\overset{p=3}{\underset{p=1}{S}}\left(\dfrac{C_{pq}}{c_{pq}}C_{p\cdot}\right)$	$\overset{p=3}{\underset{p=1}{S}}\left(\dfrac{C_{p1}}{c_{p1}}C_{p\cdot}\right)=1\cdot 1930{,}5669$	$\overset{p=3}{\underset{p=1}{S}}\left(\dfrac{C_{p2}}{c_{p2}}C_{p\cdot}\right)=1\cdot 1636{,}4101$	$\overset{p=3}{\underset{p=1}{S}}\left(\dfrac{C_{p3}}{c_{p3}}C_{p\cdot}\right)=1\cdot 2280{,}4222$	$\Sigma\left(\dfrac{C_{pq}\,C_{p\cdot}\,C_{\cdot q}}{c_{pq}}\right)=1\cdot 417{,}0530$
Same $\times\, C_{\cdot q}$	·338,2928	·596,9338	·481,8264	

Finally, by (lxxii),

$$f_1=4\times 1\cdot 636{,}6847-3\times 1\cdot 426{,}6992-3\times 1\cdot 432{,}2344+2\times 1\cdot 417{,}0530$$
$$=\cdot 804{,}0440.$$

Curves of this form are known to be accurate when we are sampling from a population, wherein there is no correlation of character and the distribution is of normal type.

Equations (lviii) to find p_1 and p_2 will then apply and we may write them

$$p_1+1=\bar{\eta}^2\left(\frac{\bar{\eta}^2(1-\bar{\eta}^2)}{\sigma^2_{\eta^2}}-1\right),\quad p_2+1=(1-\bar{\eta}^2)\left(\frac{\bar{\eta}^2(1-\bar{\eta}^2)}{\sigma^2_{\eta^2}}-1\right),$$

$$p_1+1=\bar{R}^2\left(\frac{\bar{R}^2(1-\bar{R}^2)}{\sigma^2_{R^2}}-1\right),\quad p_2+1=(1-\bar{R}^2)\left(\frac{\bar{R}^2(1-\bar{R}^2)}{\sigma^2_{R^2}}-1\right)\ \ldots\text{(lxxiv)}.$$

The values of p_1 and p_2 can be determined if the mean and variance of η^2 or R^2 be known.

If either p_1 or p_2 be large, say p_2, we are thrown back on the Type III curve

$$\left.\begin{array}{ll} & y=y_0'\left(\tfrac{1}{2}2p_2\eta^2\right)^{p_1}e^{-\frac{1}{2}2p_2\eta^2} \\ \text{or} & y=y_0'\left(\tfrac{1}{2}2p_2R^2\right)^{p_1}e^{-\frac{1}{2}2p_2R^2}\end{array}\right\}\ \ldots\ldots\ldots\ldots\text{(lxxv)},$$

and accordingly the distribution of the difference of the η^2 of two samples, or the R^2, will be given by

$$\left.\begin{array}{ll} & y=\tfrac{1}{2}MT_{p_1+\frac{1}{2}}\{p_2(\eta'^2-\eta^2)\} \\ \text{or} & y=\tfrac{1}{2}MT_{p_1+\frac{1}{2}}\{p_2(R'^2-R^2)\}\end{array}\right\}\ \ldots\ldots\ldots\ldots\text{(lxxvi)},$$

and Table I can be applied, or

$$\left.\begin{array}{ll} & P_{\eta'^2-\eta^2}=1-2S_{p_1+\frac{1}{2}}\{p_2(\eta'^2-\eta^2)\} \\ \text{and} & P_{R'^2-R^2}=1-2S_{p_1+\frac{1}{2}}\{p_2(R'^2-R^2)\}\end{array}\right\}\ \ldots\ldots\ldots\ldots\text{(lxxvii)}.$$

It is needless to add that the ratio test can be also used in such cases, or

$$\left.\begin{array}{ll} & Q_{\eta'^2/\eta^2}=2I_{\frac{1}{1+\eta'^2/\eta^2}}(p_1+1,\,p_1+1) \\ \text{and} & Q_{R'^2/R^2}=2I_{\frac{1}{1+R'^2/R^2}}(p_1+1,\,p_1+1)\end{array}\right\}\ \text{.(lxxviii)}.$$

Here as before we need *apparently* only p_1 to find Q, but we have actually to find p_2, or we have no justification for replacing (lxxiii) by (lxxv)

Illustration. In the special case of a normal population with uncorrelated characters, we know that

$$p_1=\tfrac{1}{2}(n-3),\quad p_2=\tfrac{1}{2}(N-n-2)\ \ldots\ldots\ldots\ldots\text{(lxxix)},$$

where for η^2, N is the size of the sample and n the number of arrays on which η^2 is based, while for R^2, N is again the size of the sample and n = total number of variates considered, i.e. the dependent variate and $n-1$ variates with which it is multiply correlated.

Now it is obvious that in a very large number of cases N will be large, often very large as compared with n. In such cases (lxxv) will apply, and we may write

$$\left.\begin{array}{l} P_{\eta'^2-\eta^2}=1-2S_{\frac{1}{2}(n-2)}\{\tfrac{1}{2}(N-n-2)(\eta'^2-\eta^2)\} \\ P_{R'^2-R^2}=1-2S_{\frac{1}{2}(n-2)}\{\tfrac{1}{2}(N-n-2)(R'^2-R^2)\}\end{array}\right\}\ \ldots\ldots\ldots\text{(lxxx)},$$

and for the case of the ratio,

$$\left.\begin{array}{l} Q_{\eta'^2/\eta^2}=2I_{\frac{1}{1+\eta'^2/\eta^2}}\{\tfrac{1}{2}(n-1),\,\tfrac{1}{2}(n-1)\} \\ Q_{R'^2/R^2}=2I_{\frac{1}{1+R'^2/R^2}}\{\tfrac{1}{2}(n-1),\,\tfrac{1}{2}(n-1)\}\end{array}\right\}\ \ldots\ldots\ldots\ \text{(lxxx \textit{bis})}.$$

The length which this paper has already reached hinders our providing special numerical examples for this section, but they would only be similar in type to those of the earlier sections, and accordingly little harm will be done by their omission.

(21) A limitation which detracts somewhat from a full use of the present methods must have struck the reader. Namely, we may need to compare the constants of samples which are not of the same size.

In this case our two original equations are of the form

$$y_u = \frac{M\gamma_1^{\tau_1+1}}{\Gamma(\tau_1+1)} u^{\tau_1} e^{-\gamma_1 u} \text{ and } y_v = \frac{M\gamma_2^{\tau_2+1}}{\Gamma(\tau_2+1)} v^{\tau_2} e^{-\gamma_2 v},$$

and the combined frequency surface is given by

$$w = M\frac{y_u}{M} \times \frac{y_v}{M} = \frac{M\gamma_1^{\tau_1+1}\gamma_2^{\tau_2+1}}{\Gamma(\tau_1+1)\Gamma(\tau_2+1)} e^{-(\gamma_1 u+\gamma_2 v)} u^{\tau_1} v^{\tau_2} \quad [du\,dv] \text{ ...(lxxxi).}$$

Let us first approach this from the standpoint of the ratio of v to u, and write $z = v/u$. Then (lxxxi) becomes

$$w = \frac{M\gamma_1^{\tau_1+1}\gamma_2^{\tau_2+1}}{\Gamma(\tau_1+1)\Gamma(\tau_2+1)} e^{-(\gamma_1+\gamma_2 z)u} u^{\tau_1+\tau_2+1} z^{\tau_2} \quad [du\,dz].$$

Write $\xi = (\gamma_1+\gamma_2 z)\,u$ and we have

$$w = \frac{M\gamma_1^{\tau_1+1}\gamma_2^{\tau_2+1}}{\Gamma(\tau_1+1)\Gamma(\tau_2+1)} e^{-\xi}\xi^{\tau_1+\tau_2+1} \frac{z^{\tau_2}}{(\gamma_1+\gamma_2 z)^{\tau_1+\tau_2+2}} \quad [d\xi\,dz].$$

Keeping z constant integrate out for ξ, the limits of which will be from 0 to ∞ corresponding to u from 0 to ∞. Thus we find

$$w = \frac{M\gamma_1^{\tau_1+1}\gamma_2^{\tau_2+1}}{B(\tau_1+1, \tau_2+1)} \frac{z^{\tau_2}}{(\gamma_1+\gamma_2 z)^{\tau_1+\tau_2+2}} \quad [dz] \text{ ....(lxxxii)}$$

for the distribution curves of the ratio.

Now put $z' = z\gamma_2/\gamma_1$ and integrate for the probability P'_{z_0} of a ratio greater than z_0,

$$P'_{z_0} = \frac{\gamma_1^{\tau_1+1}\gamma_2^{\tau_2+1}}{B(\tau_1+1, \tau_2+1)} \int_{z_0}^{\infty} \frac{z^{\tau_2}\,dz}{(\gamma_1+\gamma_2 z)^{\tau_1+\tau_2+2}} \text{ ....(lxxxiii).}$$

Now put $\dfrac{1}{y} = 1 + z\dfrac{\gamma_2}{\gamma_1}$, and hence $dz = -\dfrac{\gamma_1}{\gamma_2}\dfrac{1}{y^2}dy$; further when $z = \infty$, $y = 0$, and when $z = z_0$, $y = \dfrac{1}{1+\dfrac{z_0\gamma_2}{\gamma_1}}$. Thus we reach

$$P'_{z_0} = \frac{1}{B(\tau_1+1, \tau_2+1)} \int_0^{\frac{1}{1+\frac{z_0\gamma_2}{\gamma_1}}} y^{\tau_1}(1-y)^{\tau_2}\,dy$$

$$= I_{\frac{1}{1+\frac{z_0\gamma_2}{\gamma_1}}}(\tau_1+1, \tau_2+1) \text{ ....(lxxxiv),}$$

where I is the function tabled in the *Incomplete* B-*function Table* now at press.

Just as earlier (see p. 306) we should add to this the other equal wedge for $U/V > z_0$, or z taking values from 0 to $\frac{1}{z_0}$. Calling this P''_{z_0}, we have, from (lxxxiii),

$$P''_{z_0} = \frac{\gamma_1^{\tau_1+1}\gamma_2^{\tau_2+1}}{B(\tau_1+1, \tau_2+1)}\int_0^{\frac{1}{z_0}} \frac{z^{\tau_2}\,dz}{(\gamma_1+\gamma_2 z)^{\tau_1+\tau_2+2}}.$$

Now transform our integral by taking $y = \frac{z\gamma_2}{\gamma_1 + z\gamma_2}$, or limits will be $y = 0$ and $y = \frac{1}{1+\frac{z_0\gamma_1}{\gamma_2}}$, and the integral becomes

$$P''_{z_0} = \frac{1}{B(\tau_1+1, \tau_2+1)}\int_0^{\frac{1}{1+\frac{z_0\gamma_1}{\gamma_2}}} y^{\tau_2}(1-y)^{\tau_1}\,dy$$

$$= I_{\frac{1}{1+\frac{z_0\gamma_1}{\gamma_2}}}(\tau_2+1, \tau_1+1) \quad \text{...(lxxxv).}$$

Combining (lxxxiv) and (lxxxv), we reach finally

$$Q_{z_0} = I_{\frac{1}{1+\frac{z_0\gamma_2}{\gamma_1}}}(\tau_1+1, \tau_2+1) + I_{\frac{1}{1+\frac{z_0\gamma_1}{\gamma_2}}}(\tau_2+1, \tau_1+1) \quad \text{......(lxxxvi),}$$

a result involving a double entry into the B-function Table.

Thus, if we use the ratio test, (lxxxvi) shows that the determination of Q_{z_0} involves very little more trouble than in the cases we have already dealt with when $\gamma_1 = \gamma_2$ and $\tau_1 = \tau_2$.

Illustration. Suppose we have two samples rendering variances with values σ_1^2 and σ_2^2, but that the size of one sample is n_1 and of the other n_2, and we wish to test whether they may be considered as drawn from the same parent population of variance Σ^2. Then, by (xxviii),

$$\tau_1 = \tfrac{1}{2}(n_1-3),\quad \tau_2 = \tfrac{1}{2}(n_2-3),\quad \gamma_1 = \frac{n_1}{2\Sigma^2},\quad \gamma_2 = \frac{n_2}{2\Sigma^2}.$$

Accordingly

$$Q_{\sigma_2^2/\sigma_1^2} = I_{\frac{1}{1+\frac{\sigma_2^2 n_2}{\sigma_1^2 n_1}}}\{\tfrac{1}{2}(n_1-1), \tfrac{1}{2}(n_2-1)\} + I_{\frac{1}{1+\frac{\sigma_2^2 n_1}{\sigma_1^2 n_2}}}\{\tfrac{1}{2}(n_2-1), \tfrac{1}{2}(n_1-1)\} \text{...(lxxxvii).}$$

A question may arise as to whether the sum of the two I-functions may not be greater than unity and so $Q_{\sigma_2^2/\sigma_1^2} > 1$ and impossible. Clearly the denominators of both B-function ratios are equal, since

$$B\{\tfrac{1}{2}(n_1-1), \tfrac{1}{2}(n_2-1)\} = B\{\tfrac{1}{2}(n_2-1)\, \tfrac{1}{2}(n_1-1)\}.$$

We can therefore write $Q_{\sigma_2^2/\sigma_1^2}$ in the form

$$Q_{\sigma_2^2/\sigma_1^2} = \frac{\int_0^{\lambda} x^{\frac{1}{2}(n_1-3)}(1-x)^{\frac{1}{2}(n_2-3)}\,dx + \int_0^{\lambda'} x^{\frac{1}{2}(n_2-3)}(1-x)^{\frac{1}{2}(n_1-3)}\,dx}{B\{\frac{1}{2}(n_1-3), \frac{1}{2}(n_2-3)\}},$$

where $$\lambda = \frac{1}{1 + \frac{\sigma_2^2}{\sigma_1^2}\frac{n_2}{n_1}} \quad \text{and} \quad \lambda' = \frac{1}{1 + \frac{\sigma_2^2}{\sigma_1^2}\frac{n_1}{n_2}}.$$

Write $x = 1 - x'$ in the second integral, and we have

$$Q_{\sigma_2^2/\sigma_1^2} = \frac{\int_0^\lambda x^{\frac{1}{2}(n_1-3)}(1-x)^{\frac{1}{2}(n_2-3)}\,dx + \int_{1-\lambda'}^1 x'^{\frac{1}{2}(n_1-3)}(1-x')^{\frac{1}{2}(n_2-3)}\,dx'}{\mathrm{B}\{\frac{1}{2}(n_1-3), \frac{1}{2}(n_2-3)\}}.$$

Now the numerator will certainly be less than the denominator if $\lambda < 1 - \lambda'$, for then the two integrals do not make up the complete B-function. Hence our condition is that

$$\frac{1}{1 + \frac{\sigma_2^2}{\sigma_1^2}\frac{n_2}{n_1}} < \frac{\frac{\sigma_2^2}{\sigma_1^2}\frac{n_1}{n_2}}{1 + \frac{\sigma_2^2}{\sigma_1^2}\frac{n_1}{n_2}},$$

or that $$1 < \left(\frac{\sigma_2^2}{\sigma_1^2}\right)^2.$$

This is a condition always satisfied, since we have started by supposing $\sigma_2^2 > \sigma_1^2$ and asked how much greater that ratio may be than unity, without being too improbable for both samples to have a common source.

We now turn to the problem of the difference $v - u$. Taking Equation (lxxxi), let us put

$$u + v = X, \qquad v - u = Y,$$

then the surface becomes

$$w = w_0' \, e^{-\frac{1}{2}(\gamma_2-\gamma_1)Y} e^{-\frac{1}{2}(\gamma_1+\gamma_2)X} (X - Y)^{\tau_1} (X + Y)^{\tau_2} \quad [dX\,dY].$$

Now we desire to integrate this with regard to X from Y to ∞, keeping Y constant, for the portion of our surface, precisely as in § (3), p. 295. Now write $X = Yt$ and the limits of t will be 1 to ∞; thus the frequency surface for Y, which gives the difference, is

$$z = w_0'' \, e^{-\frac{1}{2}(\gamma_2-\gamma_1)Y} \, Y^{\tau_1+\tau_2+1} \int_1^\infty e^{-\frac{1}{2}(\gamma_1+\gamma_2)Yt} (t-1)^{\tau_1} (t+1)^{\tau_2}\,dt \quad \text{...(lxxxviii)}.$$

This curve is allied to the $T_m(x)$ Bessel Function curve and passes into it when $\tau_1 = \tau_2 = \tau$, but as far as we are aware has not yet been studied. Its consideration is left for another occasion. If $\tau_1 = \tau_2 = \tau$ and we write

$$Y' = \tfrac{1}{2}(\gamma_1 + \gamma_2)\,Y, \quad \rho = \frac{\gamma_2 - \gamma_1}{\gamma_2 + \gamma_1} \quad \text{and} \quad m = \tau + \tfrac{1}{2},$$

we have, by the proper choice of w_0'',

$$z = M(1 - \rho^2)^{m+\frac{1}{2}} \, e^{-\rho Y'} \, T_m(Y') \quad [dY'] \quad \text{......(lxxxviii\,\textit{bis})},$$

$T_m(Y')$ not changing sign with Y', and when we put Y' negative we get the other section of the area we need to integrate to get rid of X as in § (3), p. 295. This curve has been fully discussed in *Biometrika*, Vol. XXI. pp. 168—187. Its area when integrated for Y' between $-\infty$ and $+\infty$ equals M, and we have its ordinates tabled*.

* See *Biometrika*, Vol. XXI. pp. 195—201, or *Tables for Statisticians and Biometricians*, Part II, pp. lxxix—lxxxviii and pp. 138—144.

In terms of the difference Y,

$$z = \tfrac{1}{2}M \frac{(4\gamma_1\gamma_2)^{\tau+1}}{(\gamma_1+\gamma_2)^{2\tau+1}} e^{-\frac{1}{2}(\gamma_2-\gamma_1)Y} T_{\tau+\frac{1}{2}}\{\tfrac{1}{2}(\gamma_1+\gamma_2)Y\} \quad [dY] \ldots\text{(lxxxix)}.$$

This curve for practical purposes may be replaced by a Pearson curve of the same first four moments after $\tau = 11$*.

Illustration. It may be asked for what type of variates would it be adequate to have $\tau_1 = \tau_2$? We reply at once for η^2 and R^2, when the sizes of the samples are considerable as compared with the number of arrays in η^2 or the number of variates involved in R^2. This is the case where we are justified in using Equations (lxxv), and consequently (lxxxix). If n be the number of arrays or the total number of variates, and N, N' be the sizes of the samples, then we have

$$\tau_1 = \tau_2 = \tau = \tfrac{1}{2}(n-3),$$

while

$$\gamma_1 = \tfrac{1}{2}(N-n-2), \quad \gamma_2 = \tfrac{1}{2}(N'-n-2),$$

and the curve of distribution of, say, $\eta'^2 - \eta^2$ will be given by

$$z = \tfrac{1}{2}M \frac{\{(N-n-2)(N'-n-2)\}^{\frac{1}{2}(n-1)}}{\{\frac{1}{2}(N+N')-n-2\}^{n-2}} e^{-\frac{1}{2}(N'-N)Y} T_{\frac{1}{2}(n-2)}\{\tfrac{1}{2}(N+N'-2n-4)Y\}[dY],$$

where

$$Y = \eta'^2 - \eta^2 \quad \ldots\ldots\ldots\ldots\text{(xc)},$$

and we assume the parent population to be without association in its characters.

There is small difficulty in plotting this curve from Dr Elderton's Table of ordinates, and at present the planimeter or a quadrature formula must be applied to determine the requisite areas beyond the values $\pm(\eta'^2 - \eta^2)$ observed. An important point arises from both the Equations (lxxxviii) and (lxxxix), namely that when the sizes of the samples are unequal, then the difference of two statistical measures does not give a symmetrical curve, but one which like that of the distribution of the first product-moment coefficient may be notably skew. The fuller discussion of these curves must, however, be left to a further paper. It may be asked: Why, if the ratio-test gives an adequately simple expression for the probability, should we deal further with the more complicated expressions for the probability of the difference? The answer, we think, is that the results of the two tests are often not so closely in accord that we can be confident one may not for a particular case be more correct than the other. The probability deduced leads at once to a frequency, and the touchstone of the more correct test is that it should give this frequency the more exactly and more often. The only way of determining this is the experimental one. This would not be hard in the case of the range of χ^2's based on pairs of samples taken from a uni-variate population. It would, however, be far more laborious in the case of sample contingency tables taken from a bi-variate table. Still it is to be hoped that such work will eventually be undertaken.

(22) *General Conclusions.*

The main purpose of the present paper has been to discuss an alternative to the ratio method in dealing with a number of statistical coefficients, such as χ^2, σ^2, $\bar{x}^2$, η^2, R^2, ϕ^2, which on certain hypotheses as to the parent population obey accurately

* *loc. cit.* p. 181.

or approximately the Type III form of distribution. We have seen that the difference of two like statistical coefficients follows in its distribution the $T_m(Y)$ Bessel-Function curve, or in certain cases the more general skew form $e^{-\rho Y}\,T_m(Y)$. We have provided a table (Table I) of the probability integral of the former curve, by aid of which the probability of a given difference can be rapidly found. We have postponed discussion of a still more general curve, i.e. (lxxxviii), to a later paper.

A number of suitable illustrations were chosen by one of our number at random; that is to say without knowledge of what would flow from them, and the difference and the ratio methods both applied. The results show that in the great majority of cases the difference test is more "stringent" than the ratio test. By this we merely understand that the probability of the observed result is less by the former test. Under such conditions, however, it is reasonable to suppose that preference ought to be given to it. At any rate it justifies the use of the difference test alongside the ratio test. In order to simplify the application of the latter test, it has been approached by a new method and a table (Table II) is provided by which the probability on the ratio test can be at once determined. We have further prepared a diagram by aid of which it can be rapidly ascertained which of the tests will be found the more stringent in a particular case.

In the course of our work we have pointed out difficulties which occur in using the χ^2 methods, and warned the student of difficulties which may arise if the two series from which χ^2 is obtained are of very different sizes in the case when their relative sizes are wholly arbitrary; we indeed question whether, when the relative sizes are not "naturally" fixed, we ought to use χ^2 at all*. In the case of the comparison of two χ^2's we have with some diffidence suggested a possible method of overcoming some of the difficulty. The reader will see that much yet remains to be done—especially experimental spade-work—to obtain satisfactory tests for either the difference or ratio of these statistical coefficients in the case of different sized samples, especially small ones, when the parent population is unknown.

* There is another point about the usual χ^2 distribution frequently overlooked. Given two series of v categories each and sizes N and N', then the maximum possible value of χ^2 is $N+N'$. Accordingly the distribution curve is limited, and should take some such form as

$$y=y_0\,(\tfrac{1}{2}\chi^2)^{p_1}\,(N+N'-\chi^2)^{p_2} \qquad \text{.......}(\epsilon_1)$$

rather than

$$y=y_0\,(\tfrac{1}{2}\chi^2)^{\frac{1}{2}(v-3)}\,e^{-\frac{1}{2}\chi^2} \qquad \text{.......}(\epsilon_2).$$

If $N+N'$ be large, and then only, approach will be made to

$$p_1=\tfrac{1}{2}(\kappa-1)(\lambda-1)-1 \quad\text{and}\quad p_2=\tfrac{1}{2}(N+N')(\lambda-1)-1,$$

but in our case $\lambda=2$ and $\kappa=v$. Thus (ϵ_1) becomes

$$y=y_0'\,(\tfrac{1}{2}\chi^2)^{\frac{1}{2}(v-3)}\left(1-\frac{\frac{1}{2}\chi^2}{\frac{1}{2}(N+N')}\right)^{\frac{1}{2}(N+N')-1},$$

which takes the form of (ϵ_2) if $N+N'$ be very large. The Type III curve is therefore no more exact in the case of χ^2 than it is in the case of η^2 or R^2, it depends upon the large size of $N+N'$. This point is frequently overlooked, but it is actually involved when we replace the binomial distributions by normal curves in our deduction of (ϵ_2).

TABLE I. *Values of the Probability Integral of the $T_m(x)$ Curve, $m = 0$ to $11{\cdot}5$, $x = 0{\cdot}0$ to $18{\cdot}0$, for determining the Probability of a Difference, e.g. $\frac{1}{2}\chi'^2 - \frac{1}{2}\chi^2$. Here $v = n = 2m + 2$, or $m = \frac{1}{2}(v - 2)$.*

x	$n=2$ $m=0$	$n=3$ $m=0{\cdot}5$	$n=4$ $m=1{\cdot}0$	$n=5$ $m=1{\cdot}5$	$n=6$ $m=2{\cdot}0$	$n=7$ $m=2{\cdot}5$	$n=8$ $m=3{\cdot}0$	$n=9$ $m=3{\cdot}5$	$n=10$ $m=4{\cdot}0$	$n=11$ $m=4{\cdot}5$	$n=12$ $m=5{\cdot}0$	$n=13$ $m=5{\cdot}5$	x
0·0	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	0·0
0·1	·108 914	·047 581	·031 658	·024 960	·021 203	·018 740	·016 969	·015 620	·014 547	·013 669	·012 932	·012 302	0·1
0·2	·174 153	·090 635−	·062 572	·049 698	·042 302	·037 417	·033 897	·031 208	·029 070	·027 318	·025 847	·024 591	0·2
0·3	·223 442	·129 591	·092 382	·074 030	·063 199	·055 972	·050 740	·046 735−	·043 545+	·040 928	·038 731	·036 853	0·3
0·4	·262 795−	·164 840	·120 888	·097 808	·083 805+	·074 347	·067 459	·062 169	·057 948	·054 480	·051 566	·049 073	0·4
0·5	·295 106	·196 735−	·147 980	·120 918	·104 042	·092 487	·084 014	·077 482	·072 256	·067 956	·064 338	·061 240	0·5
0·6	·322 100	·225 594	·173 603	·143 272	·123 839	·110 344	·100 370	·092 644	·086 446	·081 335+	·077 030	·073 340	0·6
0·7	·344 920	·251 707	·197 745−	·164 805−	·143 140	·127 871	·116 490	·107 630	·100 497	·094 603	·089 628	·085 359	0·7
0·8	·364 381	·275 336	·220 416	·185 470	·161 896	·145 030	·132 345+	·122 114	·114 389	·107 740	·102 118	·097 287	0·8
0·9	·381 086	·296 715+	·241 648	·205 237	·180 067	·161 785−	·147 904	·136 971	·128 101	·120 730	·114 485−	·109 109	0·9
1·0	·395 503	·316 060	·261 487	·224 090	·197 623	·178 105+	·163 142	·151 281	·141 615−	·133 557	·126 716	·120 816	1·0
1·1	·407 998	·333 564	·279 986	·242 025−	·214 540	·193 966	·178 036	·165 322	·154 913	·146 208	·138 797	·132 395+	1·1
1·2	·418 866	·349 403	·297 205+	·259 045−	·230 804	·209 348	·192 564	·179 078	·167 981	·158 668	·150 718	·143 836	1·2
1·3	·428 348	·363 734	·313 208	·275 161	·246 405−	·224 232	·206 711	·192 530	·180 803	·170 923	·162 467	·155 129	1·3
1·4	·436 640	·376 702	·328 059	·290 393	·261 338	·238 608	·220 461	·205 666	·193 367	·182 963	·174 032	·166 264	1·4
1·5	·443 909	·388 435−	·341 825−	·304 761	·275 603	·252 465+	·233 802	·218 473	·205 658	·194 776	·185 403	·177 232	1·5
1·6	·450 294	·399 052	·354 569	·318 293	·289 207	·265 800	·246 725+	·230 939	·217 669	·206 352	·196 572	·188 025−	1·6
1·7	·455 911	·408 658	·366 356	·331 018	·302 158	·278 611	·259 224	·243 058	·229 390	·217 682	·207 530	·198 634	1·7
1·8	·460 860	·417 351	·377 248	·342 966	·314 467	·290 897	·271 294	·254 821	·240 812	·228 758	·218 271	·209 053	1·8
1·9	·465 227	·425 216	·387 303	·354 171	·326 148	·302 663	·282 932	·266 223	·251 929	·239 574	·228 785−	·219 275+	1·9
2·0	·469 085+	·432 332	·396 578	·364 665−	·337 218	·313 914	·294 138	·277 261	·262 737	·250 123	·239 067	·229 294	2·0
2·1	·472 498	·438 772	·405 129	·374 482	·347 694	·324 658	·304 913	·287 932	·273 231	·260 401	·249 114	·239 105−	2·1
2·2	·475 519	·444 598	·413 005+	·383 657	·357 596	·334 903	·315 261	·298 237	·283 408	·270 404	·258 919	·248 703	2·2
2·3	·478 195+	·449 871	·420 256	·392 222	·366 944	·344 662	·325 185+	·308 175−	·293 266	·280 129	·268 480	·258 084	2·3
2·4	·480 570	·454 641	·426 927	·400 210	·375 759	·353 944	·334 692	·317 748	·302 805+	·289 573	·277 792	·267 244	2·4
2·5	·482 677	·458 958	·433 062	·407 654	·384 062	·362 764	·343 789	·326 959	·312 026	·298 735+	·286 855+	·276 182	2·5
2·6	·484 548	·462 863	·438 701	·414 585+	·391 875+	·371 135+	·352 483	·335 812	·320 929	·307 616	·295 666	·284 894	2·6
2·7	·486 212	·466 397	·443 880	·421 033	·399 220	·379 072	·360 783	·344 312	·329 516	·316 214	·304 224	·293 378	2·7
2·8	·487 692	·469 595−	·448 636	·427 028	·406 118	·386 589	·368 698	·352 465−	·337 791	·324 532	·312 530	·301 635−	2·8
2·9	·489 008	·472 488	·453 002	·432 597	·412 591	·393 702	·376 240	·360 276	·345 757	·332 570	·320 583	·309 662	2·9
3·0	·490 181	·475 106	·457 007	·437 766	·418 660	·400 426	·383 417	·367 753	·353 418	·340 331	·328 384	·317 460	3·0

3·1	·491 225+	·477 475−	·460 680	·442 562	·424 346	·406 776	·390 243	·374 904	·360 779	·347 818	·335 935−	·325 030	3·1
3·2	·492 156	·479 619	·464 048	·447 009	·429 668	·412 769	·396 727	·381 735+	·367 846	·355 035−	·343 237	·332 372	3·2
3·3	·492 986	·481 558	·467 134	·451 130	·434 646	·418 419	·402 883	·388 257	·374 624	·361 984	·350 294	·339 487	3·3
3·4	·493 726	·483 313	·469 962	·454 946	·439 299	·423 742	·408 720	·394 477	·381 119	·368 671	·357 107	·346 378	3·4
3·5	·494 387	·484 901	·472 552	·458 479	·443 646	·428 753	·414 252	·400 404	·387 338	·375 099	·363 680	·353 045−	3·5
3·6	·494 977	·486 338	·474 923	·461 747	·447 703	·433 467	·419 489	·406 047	·393 288	·381 275−	·370 016	·359 492	3·6
3·7	·495 503	·487 638	·477 094	·464 769	·451 489	·437 898	·424 445+	·411 417	·398 976	·387 202	·376 120	·365 721	3·7
3·8	·495 974	·488 815−	·479 081	·467 562	·455 018	·442 060	·429 130	·416 522	·404 410	·392 888	·381 996	·371 735+	3·8
3·9	·496 394	·489 879	·480 899	·470 143	·458 306	·445 967	·433 557	·421 371	·409 596	·398 338	·387 647	·377 538	3·9
4·0	·496 770	·490 842	·482 561	·472 527	·461 369	·449 632	·437 735+	·425 974	·414 542	·403 557	·393 079	·383 133	4·0
4·5	·498 132	·494 446	·488 965−	·481 948	·473 757	·464 764	·455 298	·445 627	·435 951	·426 416	·417 120	·408 125+	4·5
5·0	·498 915−	·496 631	·493 040	·488 209	·482 312	·475 575−	·468 229	·460 485−	·452 519	·444 473	·436 457	·428 550−	5·0
5·5	·499 367	·497 957	·495 623	·492 337	·488 159	·483 206	·477 622	·471 558	·465 150+	·458 523	·451 776	·444 994	5·5
6·0	·499 630	·498 761	·497 254	·495 042	·492 121	·488 536	·484 366	·479 705+	·474 654	·469 306	·463 748	·458 053	6·0
6·5	·499 783	·499 248	·498 281	·496 805+	·494 785+	·492 224	·489 156	·485 633	·481 719	·477 482	·472 986	·468 294	6·5
7·0	·499 873	·499 544	·498 926	·497 948	·496 565−	·494 757	·492 528	·489 903	·486 916	·483 611	·480 035−	·476 234	7·0
7·5	·499 925+	·499 723	·499 330	·498 686	·497 747	·496 483	·494 883	·492 950+	·490 701	·488 160	·485 357	·482 323	7·5
8·0	·499 956	·499 832	·499 583	·499 161	·498 528	·497 652	·496 514	·495 108	·493 435−	·491 505+	·489 336	·486 948	8·0
8·5	·499 974	·499 898	·499 740	·499 466	·499 041	·498 439	·497 638	·496 624	·495 392	·493 944	·492 286	·490 429	8·5
9·0	·499 985−	·499 938	·499 839	·499 661	·499 378	·498 966	·498 406	·497 682	·496 784	·495 708	·494 454	·493 026	9·0
9·5	·499 991	·499 963	·499 900	·499 785−	·499 597	·499 318	·498 929	·498 416	·497 767	·496 975+	·496 036	·494 949	9·5
10·0	·499 995−	·499 977	·499 938	·499 864	·499 740	·499 552	·499 284	·498 923	·498 457	·497 879	·497 182	·496 363	10·0
10·5	·499 997	·499 986	·499 962	·499 914	·499 833	·499 706	·499 522	·499 270	·498 939	·498 520	·498 007	·497 395−	10·5
11·0	·499 998	·499 992	·499 976	·499 946	·499 892	·499 808	·499 683	·499 507	·499 273	·498 972	·498 597	·498 143	11·0
11·5	·499 999	·499 995−	·499 985+	·499 966	·499 931	·499 875−	·499 790	·499 669	·499 504	·499 289	·499 017	·498 683	11·5
12·0	·499 999	·499 997	·499 991	·499 978	·499 956	·499 919	·499 861	·499 778	·499 663	·499 510	·499 315−	·499 070	12·0
12·5	·500 000−	·499 998	·499 994	·499 986	·499 972	·499 947	·499 909	·499 852	·499 772	·499 664	·499 524	·499 346	12·5
13·0	·500 000−	·499 999	·499 997	·499 992	·499 982	·499 966	·499 940	·499 901	·499 846	·499 770	·499 670	·499 542	13·0
13·5	·500 000−	·499 999	·499 998	·499 995−	·499 989	·499 978	·499 961	·499 934	·499 896	·499 844	·499 773	·499 681	13·5
14·0	·500 000−	·500 000−	·499 999	·499 997	·499 993	·499 986	·499 974	·499 957	·499 930	·499 894	·499 844	·499 [illegible]78	14·0
14·5	·500 000−	·500 000−	·499 999	·499 998	·499 995+	·499 991	·499 983	·499 971	·499 953	·499 928	·499 893	·499 846	14·5
15·0	·500 000−	·500 000−	·499 999	·499 999	·499 997	·499 994	·499 989	·499 981	·499 969	·499 951	·499 927	·499 894	15·0
15·5	·500 000−	·500 000−	·500 000−	·499 999	·499 998	·499 996	·499 993	·499 988	·499 979	·499 967	·499 950	·499 927	15·5
16·0	·500 000−	·500 000−	·500 000−	·499 999	·499 999	·499 998	·499 995+	·499 992	·499 986	·499 978	·499 966	·499 950−	16·0
16·5	·500 000−	·500 000−	·500 000−	·500 000−	·499 999	·499 998	·499 997	·499 995−	·499 991	·499 985+	·499 977	·499 966	16·5
17·0	·500 000−	·500 000−	·500 000−	·500 000−	·499 999	·499 999	·499 998	·499 996	·499 994	·499 990	·499 985−	·499 977	17·0
17·5	·500 000−	·500 000−	·500 000−	·500 000−	·500 000−	·499 999	·499 999	·499 998	·499 996	·499 993	·499 990	·499 984	17·5
18·0	·500 000−	·500 000−	·500 000−	·500 000−	·500 000−	·500 000−	·499 999	·499 999	·499 997	·499 996	·499 993	·499 989	18·0

TABLE I (*continued*).

x	$n=14$ $m=6{\cdot}0$	$n=15$ $m=6{\cdot}5$	$n=16$ $m=7{\cdot}0$	$n=17$ $m=7{\cdot}5$	$n=18$ $m=8{\cdot}0$	$n=19$ $m=8{\cdot}5$	$n=20$ $m=9{\cdot}0$	$n=21$ $m=9{\cdot}5$	$n=22$ $m=10{\cdot}0$	$n=23$ $m=10{\cdot}5$	$n=24$ $m=11{\cdot}0$	$n=25$ $m=11{\cdot}5$	x
0·0	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	·000 000	0·0
0·5	·058 549	·056 184	·054 083	·052 201	·050 502	·048 959	·047 549	·046 254	·045 060	·043 953	·042 924	·041 965−	0·5
1·0	·115 662	·111 112	·107 056	·103 411	·100 114	·097 112	·094 363	·091 835−	·089 499	·087 332	·085 315+	·083 432	1·0
1·5	·170 032	·163 630	·157 890	·152 708	·147 998	·143 697	·139 746	·136 102	·132 727	·129 591	·126 664	·123 928	1·5
2·0	·220 585−	·212 768	·205 705+	·199 287	·193 422	·188 039	·183 074	·178 478	·174 207	·170 226	·166 503	·163 013	2·0
2·5	·266 544	·257 797	·249 820	·242 512	·235 789	·229 580	·223 825−	·218 472	·213 478	·208 806	·204 422	·200 300	2·5
3·0	·307 450−	·298 251	·289 771	·281 932	·274 662	·267 901	·261 595+	·255 699	·250 171	·244 976	·240 084	·235 466	3·0
3·5	·343 144	·333 921	·325 320	·317 287	·309 771	·302 726	·296 110	·289 885−	·284 016	·278 474	·273 230	·268 261	3·5
4·0	·373 720	·364 826	·356 427	·348 497	·341 005−	·333 922	·327 219	·320 869	·314 846	·309 126	·303 687	·298 510	4·0
4·5	·399 468	·391 165+	·383 222	·375 635−	·368 393	·361 483	·354 890	·348 598	·342 590	·336 849	·331 361	·326 110	4·5
5·0	·420 808	·413 271	·405 963	·398 897	·392 080	·385 513	·379 192	·373 112	·367 264	·361 641	·356 233	·351 030	5·0
5·5	·438 239	·431 561	·424 996	·418 571	·412 302	·406 202	·400 277	·394 531	·388 962	·383 571	·378 352	·373 303	5·5
6·0	·452 284	·446 493	·440 721	·434 999	·429 354	·423 803	·418 360	·413 036	·407 835+	·402 764	·397 822	·393 012	6·0
6·5	·463 461	·458 534	·453 554	·448 555−	·443 564	·438 606	·433 696	·428 851	·424 080	·419 391	·414 792	·410 285+	6·5
7·0	·472 253	·468 132	·463 908	·459 614	·455 277	·450 921	·446 566	·442 228	·437 921	·433 656	·429 441	·425 284	7·0
7·5	·479 094	·475 700	·472 174	·468 542	·464 831	·461 063	·457 258	·453 432	·449 601	·445 778	·441 971	·438 191	7·5
8·0	·484 364	·481 609	·478 706	·475 678	·472 547	·469 333	·466 055−	·462 728	·459 367	·455 986	·452 594	·449 202	8·0
8·5	·488 387	·486 179	·483 819	·481 327	·478 720	·476 015+	·473 227	·470 371	·467 460	·464 507	·461 523	·458 517	8·5
9·0	·491 432	·489 682	·487 787	·485 761	·483 616	·481 366	·479 024	·476 601	·474 110	·471 562	·468 967	·466 333	9·0
9·5	·493 717	·492 346	·490 841	·489 212	·487 467	·485 617	·483 670	·481 639	·479 531	·477 356	·475 123	·472 632	9·5
10·0	·495 420	·494 356	·493 173	·491 876	·490 472	·488 966	·487 366	·485 679	·483 914	·482 077	·480 177	·478 219	10·0
10·5	·496 680	·495 862	·494 940	·493 918	·492 798	·491 585−	·490 282	·488 896	·487 432	·485 896	·484 294	·482 632	10·5
11·0	·497 606	·496 982	·496 271	·495 472	·494 588	·493 618	·492 568	·491 439	·490 237	·488 964	·487 626	·486 227	11·0
11·5	·498 282	·497 810	·497 266	·496 647	·495 954	·495 187	·494 347	·493 436	·492 456	·491 411	·490 303	·489 136	11·5
12·0	·498 773	·498 419	·498 005+	·497 530	·496 991	·496 389	·495 723	·494 993	·494 202	·493 351	·492 441	·491 475+	12·0
12·5	·499 127	·498 864	·498 552	·498 189	·497 773	·497 304	·496 780	·496 200	·495 566	·494 878	·494 137	·493 345−	12·5
13·0	·499 382	·499 187	·498 953	·498 678	·498 360	·497 997	·497 587	·497 131	·496 626	·496 074	·495 475+	·494 829	13·0
13·5	·499 564	·499 421	·499 247	·499 040	·498 798	·498 519	·498 201	·497 843	·497 445−	·497 005+	·496 524	·496 002	13·5
14·0	·499 694	·499 589	·499 460	·499 305+	·499 122	·498 909	·498 664	·498 386	·498 074	·497 726	·497 342	·496 923	14·0
14·5	·499 786	·499 709	·499 614	·499 499	·499 362	·499 200	·499 013	·498 798	·498 554	·498 281	·497 977	·497 642	14·5
15·0	·499 851	·499 795+	·499 726	·499 641	·499 538	·499 416	·499 273	·499 108	·498 919	·498 706	·498 467	·498 201	15·0
15·5	·499 896	·499 856	·499 806	·499 743	·499 667	·499 575+	·499 467	·499 341	·499 195+	·499 030	·498 843	·498 633	15·5
16·0	·499 928	·499 899	·499 863	·499 817	·499 760	·499 692	·499 611	·499 515−	·499 404	·499 276	·499 130	·498 966	16·0
16·5	·499 950+	·499 930	·499 903	·499 870	·499 828	·499 778	·499 717	·499 644	·499 559	·499 461	·499 348	·499 220	16·5
17·0	·499 966	·499 951	·499 932	·499 908	·499 878	·499 840	·499 795	·499 740	·499 676	·492 601	·499 514	·499 415−	17·0
17·5	·499 976	·499 966	·499 953	·499 935+	·499 913	·499 886	·499 852	·499 812	·499 762	·499 706	·499 639	·499 564	17·5
18·0	·499 984	·499 977	·499 967	·499 954	·499 938	·499 918	·499 893	·499 863	·499 826	·499 783	·499 732	·499 674	18·0

TABLE II. *Values of* $I_x\{\frac{1}{2}(v-1), \frac{1}{2}(v-1)\}$, $v=2$ *to* 25 *and* $x=\cdot01$ *to* $\cdot50$, *for determining the Probability of Ratios, e.g.* χ'^2/χ^2.

$x=\cdot01$—$\cdot50$ $v=2$—7

x	$v=2$ $\frac{1}{2}(v-1)=0\cdot5$	3 1	4 1·5	5 2	6 2·5	7 3	x
·01	·063 768 56	·010 000 00	·001 692 55⁺	·000 298 00	·000 053 74	·000 009 85⁺	·01
·02	·090 334 47	·020 000 00	·004 772 77	·001 184 00	·000 300 75⁻	·000 077 62	·02
·03	·110 824 69	·030 000 00	·008 741 44	·002 646 00	·000 819 78	·000 258 00	·03
·04	·128 188 43	·040 000 00	·013 417 07	·004 672 00	·001 664 48	·000 602 21	·04
·05	·143 566 29	·050 000 00	·018 693 04	·007 250 00	·002 875 76	·001 158 12	·05
·06	·157 542 43	·060 000 00	·024 496 26	·010 368 00	·004 486 12	·001 970 27	·06
·07	·170 463 43	·070 000 00	·030 772 22	·014 014 00	·006 521 83	·003 079 93	·07
·08	·182 554 89	·080 000 00	·037 477 97	·018 176 00	·009 004 21	·004 525 26	·08
·09	·193 973 37	·090 000 00	·044 578 42	·022 842 00	·011 950 56	·006 341 28	·09
·10	·204 832 76	·100 000 00	·052 044 02	·028 000 00	·015 374 72	·008 560 00	·10
·11	·215 219 03	·110 000 00	·059 849 42	·033 638 00	·019 287 59	·011 210 48	·11
·12	·225 198 90	·120 000 00	·067 972 43	·039 744 00	·023 697 45⁺	·014 318 90	·12
·13	·234 825 47	·130 000 00	·076 393 38	·046 306 00	·028 610 26	·017 908 63	·13
·14	·244 141 78	·140 000 00	·085 094 64	·053 312 00	·034 029 90	·022 000 29	·14
·15	·253 183 31	·150 000 00	·094 060 20	·060 750 00	·039 958 35⁻	·026 611 88	·15
·16	·261 979 76	·160 000 00	·103 275 47	·068 608 00	·046 395 85⁺	·031 758 75⁻	·16
·17	·270 556 26	·170 000 00	·112 726 99	·076 874 00	·053 341 09	·037 453 76	·17
·18	·278 934 34	·180 000 00	·122 402 29	·085 536 00	·060 791 27	·043 707 34	·18
·19	·287 132 59	·190 000 00	·132 289 75⁻	·094 582 00	·068 742 25⁻	·050 527 51	·19
·20	·295 167 24	·200 000 00	·142 378 49	·104 000 00	·077 188 63	·057 920 00	·20
·21	·303 052 54	·210 000 00	·152 658 26	·113 778 00	·086 123 84	·065 888 31	·21
·22	·310 801 12	·220 000 00	·163 119 39	·123 904 00	·095 540 22	·074 433 78	·22
·23	·318 424 23	·230 000 00	·173 752 67	·134 366 00	·105 429 11	·083 555 66	·23
·24	·325 931 94	·240 000 00	·184 549 36	·145 152 00	·115 780 88	·093 251 17	·24
·25	·333 333 33	·250 000 00	·195 501 11	·156 250 00	·126 585 00⁻	·103 515 62	·25
·26	·340 636 66	·260 000 00	·206 599 90	·167 648 00	·137 830 12	·114 342 43	·26
·27	·347 849 40	·270 000 00	·217 838 05⁺	·179 334 00	·149 504 09	·125 723 19	·27
·28	·354 978 44	·280 000 00	·229 208 15⁻	·191 296 00	·161 594 04	·137 647 82	·28
·29	·362 030 07	·290 000 00	·240 703 03	·203 522 00	·174 086 40	·150 104 54	·29
·30	·369 010 12	·300 000 00	·252 315 79	·216 000 00	·186 966 96	·163 080 00	·30
·31	·375 923 99	·310 000 00	·264 039 69	·228 718 00	·200 220 90	·176 559 34	·31
·32	·382 776 69	·320 000 00	·275 868 23	·241 664 00	·213 832 81	·190 526 26	·32
·33	·389 572 92	·330 000 00	·287 795 04	·254 826 00	·227 786 80	·204 963 09	·33
·34	·396 317 08	·340 000 00	·299 813 93	·268 192 00	·242 066 44	·219 850 85⁺	·34
·35	·403 013 32	·350 000 00	·311 918 83	·281 750 00	·256 654 85⁻	·235 169 38	·35
·36	·409 665 53	·360 000 00	·324 103 83	·295 488 00	·271 534 73	·250 897 31	·36
·37	·416 277 43	·370 000 00	·336 363 11	·309 394 00	·286 688 37	·267 012 22	·37
·38	·422 852 55⁺	·380 000 00	·348 690 97	·323 456 00	·302 097 72	·283 490 70	·38
·39	·429 394 26	·390 000 00	·361 081 79	·337 662 00	·317 744 35⁺	·300 308 37	·39
·40	·435 905 78	·400 000 00	·373 530 04	·352 000 00	·333 609 56	·317 440 00	·40
·41	·442 390 22	·410 000 00	·386 030 28	·366 458 00	·349 674 36	·334 859 57	·41
·42	·448 850 58	·420 000 00	·398 577 12	·381 024 00	·365 919 48	·352 540 34	·42
·43	·455 289 74	·430 000 00	·411 165 24	·395 686 00	·382 325 47	·370 454 92	·43
·44	·461 710 54	·440 000 00	·423 789 37	·410 432 00	·398 872 64	·388 575 33	·44
·45	·468 115 72	·450 000 00	·436 444 29	·425 250 00	·415 541 14	·406 873 12	·45
·46	·474 507 97	·460 000 00	·449 124 80	·440 128 00	·432 310 98	·425 319 39	·46
·47	·480 889 93	·470 000 00	·461 825 74	·455 054 00	·449 162 04	·443 884 85⁺	·47
·48	·487 264 21	·480 000 00	·474 542 00	·470 016 00	·466 074 10	·462 539 98	·48
·49	·493 633 38	·490 000 00	·487 268 45⁺	·485 002 00	·483 026 87	·481 255 00⁻	·49
·50	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·50

TABLE II (*continued*).

x = ·01—·50 v = 8—13

x	$v=8$ $\frac{1}{2}(v-1)=3{\cdot}5$	9 4	10 4·5	11 5	12 5·5	13 6	x
·01	·000 001 83	·000 000 34	·000 000 06	·000 000 01			·01
·02	·000 020 26	·000 005 34	·000 001 41	·000 000 38	·000 000 10	·000 000 03	·02
·03	·000 082 12	·000 026 36	·000 008 52	·000 002 77	·000 000 90	·000 000 29	·03
·04	·000 220 32	·000 081 28	·000 030 18	·000 011 27	·000 004 23	·000 001 59	·04
·05	·000 471 53	·000 193 58	·000 079 99	·000 033 22	·000 013 86	·000 005 80	·05
·06	·000 874 69	·000 391 49	·000 176 34	·000 079 84	·000 036 30	·000 016 56	·06
·07	·001 469 99	·000 707 24	·000 342 42	·000 166 62	·000 081 42	·000 039 92	·07
·08	·002 298 07	·001 176 28	·000 605 82	·000 313 58	·000 162 98	·000 085 00+	·08
·09	·003 399 45−	·001 836 58	·000 998 30	·000 545 31	·000 299 08	·000 164 59	·09
·10	·004 813 96	·002 728 00	·001 555 21	·000 890 92	·000 512 41	·000 295 71	·10
·11	·006 580 38	·003 891 63	·002 315 12	·001 383 83	·000 830 43	·000 499 98	·11
·12	·008 736 05+	·005 369 26	·003 319 17	·002 061 48	·001 285 31	·000 803 99	·12
·13	·011 316 60	·007 202 82	·004 610 61	·002 964 92	·001 913 91	·001 239 42	·13
·14	·014 355 68	·009 433 86	·006 234 16	·004 138 37	·002 757 43	·001 843 09	·14
·15	·017 884 79	·012 103 17	·008 235 49	·005 628 66	·003 861 14	·002 656 86	·15
·16	·021 933 07	·015 250 28	·010 660 61	·007 484 70	·005 273 88	·003 727 40	·16
·17	·026 527 17	·018 913 11	·013 555 36	·009 756 81	·007 047 55+	·005 105 78	·17
·18	·031 691 16	·023 127 64	·016 964 83	·012 496 17	·009 236 46	·006 847 03	·18
·19	·037 446 37	·027 927 56	·020 932 88	·015 754 09	·011 896 70	·009 009 50+	·19
·20	·043 811 41	·033 344 00	·025 501 63	·019 581 44	·015 085 40	·011 654 21	·20
·21	·050 802 04	·039 405 31	·030 711 00	·024 027 96	·018 859 97	·014 843 99	·21
·22	·058 431 15+	·046 136 81	·036 598 30	·029 141 65+	·023 277 34	·018 642 73	·22
·23	·066 708 79	·053 560 62	·043 197 81	·034 968 19	·028 393 20	·023 114 40	·23
·24	·075 642 09	·061 695 47	·050 540 44	·041 550 31	·034 261 18	·028 322 19	·24
·25	·085 235 33	·070 556 64	·058 653 40	·048 927 31	·040 932 12	·034 327 51	·25
·26	·095 489 93	·080 155 78	·067 559 93	·057 134 50−	·048 453 32	·041 189 04	·26
·27	·106 404 49	·090 500 89	·077 279 01	·066 202 79	·056 867 88	·048 961 83	·27
·28	·117 974 83	·101 596 24	·087 825 23	·076 158 25+	·066 213 99	·057 696 37	·28
·29	·130 194 04	·113 442 36	·099 208 53	·087 021 75+	·076 524 39	·067 437 74	·29
·30	·143 052 55+	·126 036 00	·111 434 17	·098 808 66	·087 825 79	·078 224 79	·30
·31	·156 538 20	·139 370 22	·124 502 57	·111 528 58	·100 138 40	·090 089 40	·31
·32	·170 636 31	·153 434 35+	·138 409 26	·125 185 15+	·113 475 54	·103 055 86	·32
·33	·185 329 77	·168 214 12	·153 144 92	·139 775 93	·127 843 33	·117 140 27	·33
·34	·200 599 12	·183 691 68	·168 695 35−	·155 292 26	·143 240 39	·132 350 07	·34
·35	·216 422 66	·199 845 73	·185 041 56	·171 719 29	·159 657 74	·148 683 72	·35
·36	·232 776 57	·216 651 65+	·202 159 85−	·189 035 96	·177 078 63	·166 130 40	·36
·37	·249 634 99	·234 081 58	·220 021 92	·207 215 12	·195 478 62	·184 669 86	·37
·38	·266 970 13	·252 104 62	·238 595 07	·226 223 66	·214 825 60	·204 272 46	·38
·39	·284 752 43	·270 686 94	·257 842 33	·246 022 66	·235 079 98	·224 899 18	·39
·40	·302 950 64	·289 792 00	·277 722 72	·266 567 68	·256 194 90	·246 501 87	·40
·41	·321 531 96	·309 380 70	·298 191 47	·287 809 02	·278 116 57	·269 023 54	·41
·42	·340 462 20	·329 411 61	·319 200 28	·309 692 05−	·300 784 62	·292 398 79	·42
·43	·359 705 85+	·349 841 12	·340 697 63	·332 157 58	·324 132 59	·316 554 32	·43
·44	·379 226 29	·370 623 73	·362 629 05+	·355 142 26	·348 088 40	·341 409 58	·44
·45	·398 985 85−	·391 712 20	·384 937 50−	·378 579 05+	·372 574 95+	·366 877 42	·45
·46	·418 946 02	·413 057 85+	·407 563 65−	·402 397 68	·397 510 74	·392 864 92	·46
·47	·439 067 55+	·434 610 74	·430 446 28	·426 525 10	·422 810 52	·419 274 23	·47
·48	·459 310 62	·456 319 93	·453 522 63	·450 886 10	·448 385 98	·446 003 47	·48
·49	·479 634 96	·478 133 75−	·476 728 77	·475 403 74	·474 146 52	·472 947 73	·49
·50	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·50

TABLE II (*continued*).

x = ·01—·50 v = 14—19

x	$v=14$, $\frac{1}{2}(v-1)=6{\cdot}5$	15, 7	16, 7·5	17, 8	18, 8·5	19, 9	x
·01							·01
·02	·000 000 01						·02
·03	·000 000 10	·000 000 03	·000 000 01				·03
·04	·000 000 60	·000 000 23	·000 000 09	·000 000 03	·000 000 01		·04
·05	·000 002 44	·000 001 03	·000 000 43	·000 000 18	·000 000 08	·000 000 03	·05
·06	·000 007 58	·000 003 48	·000 001 60	·000 000 74	·000 000 34	·000 000 16	·06
·07	·000 019 64	·000 009 68	·000 004 79	·000 002 37	·000 001 18	·000 000 58	·07
·08	·000 044 46	·000 [illegible] 32	·000 012 25+	·000 006 45+	·000 003 40	·000 001 80	·08
·09	·000 090 84	·000 050 26	·000 027 87	·000 015 48	·000 008 62	·000 004 80	·09
·10	·000 171 14	·000 099 29	·000 057 73	·000 033 63	·000 019 62	·000 011 47	·10
·11	·000 301 88	·000 182 71	·000 110 82	·000 067 34	·000 040 98	·000 024 98	·11
·12	·000 504 31	·000 317 09	·000 199 79	·000 126 12	·000 079 74	·000 050 49	·12
·13	·000 804 83	·000 523 85+	·000 341 67	·000 223 26	·000 146 11	·000 095 76	·13
·14	·001 235 25+	·000 829 80	·000 558 56	·000 376 66	·000 254 40	·000 172 06	·14
·15	·001 833 03	·001 267 55−	·000 878 26	·000 609 61	·000 423 79	·000 295 03	·15
·16	·002 641 24	·001 875 80	·001 334 80	·000 951 48	·000 679 29	·000 485 63	·16
·17	·003 708 45+	·002 699 49	·001 968 83	·001 438 39	·001 052 45−	·000 771 11	·17
·18	·005 088 43	·003 789 71	·002 827 82	·002 113 61	·001 582 13	·001 185 88	·18
·19	·006 839 73	·005 203 53	·003 966 12	·003 027 93	·002 315 06	·001 772 34	·19
·20	·009 025 04	·007 003 56	·005 444 77	·004 239 75−	·003 306 16	·002 581 46	·20
·21	·011 710 56	·009 257 38	·007 331 16	·005 814 91	·004 618 74	·003 673 27	·21
·22	·014 965 07	·012 036 80	·009 698 41	·007 826 37	·006 324 38	·005 116 96	·22
·23	·018 859 04	·015 416 95−	·012 624 57	·010 353 56	·008 502 51	·006 990 84	·23
·24	·023 463 59	·019 475 23	·016 191 67	·013 481 54	·011 239 78	·009 381 85+	·24
·25	·028 849 42	·024 290 14	·020 484 48	·017 299 84	·014 629 02	·012 384 78	·25
·26	·035 085 64	·029 940 04	·025 589 25−	·021 901 19	·018 768 05−	·016 101 16	·26
·27	·042 238 66	·036 501 79	·031 592 19	·027 379 97	·023 758 09	·020 637 80	·27
·28	·050 370 99	·044 049 36	·038 577 95+	·033 830 47	·029 702 04	·026 104 99	·28
·29	·059 540 16	·052 652 47	·046 627 96	·041 345 12	·036 702 44	·032 614 43	·29
·30	·069 797 56	·062 375 21	·055 818 76	·050 012 54	·044 859 37	·040 276 94	·30
·31	·081 187 50+	·073 274 66	·066 220 38	·059 915 58	·054 268 23	·049 199 94	·31
·32	·093 746 17	·085 399 64	·077 894 69	·071 129 37	·065 017 36	·059 484 84	·32
·33	·107 500 85+	·098 789 54	·090 893 88	·083 719 39	·077 185 82	·071 224 36	·33
·34	·122 469 15+	·113 473 24	·105 259 06	·097 739 68	·090 841 13	·084 499 88	·34
·35	·138 658 39	·129 468 23	·121 018 94	·113 231 13	·106 037 17	·099 378 86	·35
·36	·156 065 13	·146 779 78	·138 188 78	·130 220 07	·122 812 26	·115 912 48	·36
·37	·174 674 82	·165 400 43	·156 769 47	·148 716 95+	·141 187 52	·134 133 51	·37
·38	·194 461 65−	·185 309 55+	·176 746 88	·168 715 38	·161 165 45+	·154 054 43	·38
·39	·215 388 47	·206 473 17	·198 091 45+	·190 191 42	·182 728 93	·175 666 06	·39
·40	·237 406 98	·228 843 95+	·220 758 00	·213 103 18	·205 840 51	·198 936 49	·40
·41	·260 457 95+	·252 361 44	·244 685 83	·237 390 78	·230 442 12	·223 810 52	·41
·42	·284 471 69	·276 952 44	·269 799 10	·262 976 59	·256 455 25−	·250 209 65+	·42
·43	·309 368 62	·302 531 68	·296 007 46	·289 765 87	·283 781 46	·278 032 49	·43
·44	·335 059 97	·329 002 56	·323 206 91	·317 647 66	·312 303 36	·307 155 73	·44
·45	·361 448 66	·356 258 18	·351 280 94	·346 496 07	·341 885 97	·337 435 62	·45
·46	·388 430 23	·384 182 49	·380 101 86	·376 171 82	·372 378 46	·368 709 89	·46
·47	·415 893 90	·412 651 52	·409 532 32	·406 524 00	·403 616 22	·400 800 13	·47
·48	·443 723 76	·441 534 89	·439 427 07	·437 392 14	·435 423 24	·433 514 52	·48
·49	·471 799 96	·470 697 27	·469 634 78	·468 608 41	·467 614 74	·466 650 88	·49
·50	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·50

TABLE II (*continued*).

$x=\cdot01$—$\cdot50$ $v=20$—25

x	$v=20$ $\frac{1}{2}(v-1)=9\cdot5$	21 10	22 10·5	23 11	24 11·5	25 12	x
·01							·01
·02							·02
·03							·03
·04							·04
·05	·000 000 01	·000 000 01					·05
·06	·000 000 07	·000 000 03	·000 000 01	·000 000 01			·06
·07	·000 000 29	·000 000 14	·000 000 07	·000 000 04	·000 000 02	·000 000 01	·07
.08	·000 000 95+	·000 000 50+	·000 000 27	·000 000 14	·000 000 07	·000 000 04	·08
·09	·000 002 68	·000 001 50−	·000 000 84	·000 000 47	·000 000 26	·000 000 15−	·09
·10	·000 006 71	·000 003 93	·000 002 30	·000 001 35+	·000 000 80	·000 000 47	·10
·11	·000 015 25−	·000 009 32	·000 005 70	·000 003 49	·000 002 14	·000 001 31	·11
·12	·000 032 01	·000 020 32	·000 012 91	·000 008 21	·000 005 23	·000 003 33	·12
·13	·000 062 85−	·000 041 29	·000 027 16	·000 017 88	·000 011 78	·000 007 77	·13
·14	·000 116 53	·000 079 01	·000 053 62	·000 036 43	·000 024 77	·000 016 86	·14
·15	·000 205 65−	·000 143 51	·000 100 25−	·000 070 09	·000 049 05+	·000 034 36	·15
·16	·000 347 61	·000 249 10	·000 178 69	·000 128 30	·000 092 20	·000 066 31	·16
·17	·000 565 66	·000 415 42	·000 305 38	·000 224 70	·000 165 47	·000 121 96	·17
·18	·000 889 94	·000 668 58	·000 502 78	·000 378 44	·000 285 08	·000 214 92	·18
·19	·001 358 45+	·001 042 34	·000 800 56	·000 615 42	·000 473 48	·000 364 55+	·19
·20	·002 017 96	·001 579 12	·001 236 90	·000 969 70	·000 760 83	·000 597 39	·20
·21	·002 924 66	·002 331 03	·001 859 63	·001 484 85−	·001 186 54	·000 948 85+	·21
·22	·004 144 68	·003 360 54	·002 727 27	·002 215 21	·001 800 70	·001 464 80	·22
·23	·005 754 21	·004 741 03	·003 909 79	·003 226 96	·002 665 41	·002 203 14	·23
·24	·007 839 40	·006 556 89	·005 489 06	·004 598 86	·003 855 91	·003 235 21	·24
·25	·010 495 75+	·008 903 28	·007 558 96	·006 422 71	·005 461 25−	·004 646 85−	·25
·26	·013 827 25+	·011 885 44	·010 224 94	·008 803 24	·007 584 63	·006 539 01	·26
·27	·017 945 01	·015 617 59	·013 603 19	·011 857 54	·010 343 10	·009 027 87	·27
·28	·022 965 53	·020 221 26	·017 819 13	·015 713 86	·013 866 69	·012 244 30	·28
·29	·029 008 64	·025 823 31	·023 005 53	·020 509 80	·018 296 81	·016 332 51	·29
·30	·036 195 01	·032 553 36	·029 300 01	·026 389 94	·023 784 00	·021 448 00	·30
·31	·044 643 52	·040 540 98	·036 842 03	·033 502 81	·030 484 93	·027 754 67	·31
·32	·054 468 26	·049 912 51	·045 769 58	·041 997 35−	·038 558 78	·035 421 14	·32
·33	·065 775 56	·060 787 68	·056 215 44	·052 018 98	·048 163 08	·044 616 45+	·33
·34	·078 660 87	·073 276 06	·068 303 28	·063 705 29	·059 449 03	·055 505 05−	·34
·35	·093 205 72	·087 473 60	·082 143 66	·077 181 50+	·072 556 54	·068 241 42	·35
·36	·109 474 85−	·103 459 16	·097 830 02	·092 556 03	·087 609 19	·082 964 44	·36
·37	·127 513 53	·121 291 33	·115 434 93	·109 915 98	·104 709 14	·099 791 71	·37
·38	·147 345 27	·141 005 53	·135 006 62	·129 323 12	·123 932 38	·118 814 04	·38
·39	·168 969 90	·162 611 64	·156 565 91	·150 810 19	·145 324 36	·140 090 39	·39
·40	·192 362 12	·186 092 02	·180 103 87	·174 377 87	·168 896 34	·163 643 44	·40
·41	·217 470 63	·211 400 28	·205 580 00	·199 992 54	·194 622 51	·189 456 14	·41
·42	·244 217 84	·238 460 66	·232 921 30	·227 584 86	·222 438 11	·217 469 18	·42
·43	·272 500 17	·267 168 16	·262 022 11	·257 049 35−	·252 238 59	·247 579 79	·43
·44	·302 188 97	·297 389 37	·292 744 90	·288 244 94	·283 880 06	·279 641 85+	·44
·45	·333 132 07	·328 964 09	·324 921 83	·320 996 61	·317 180 75−	·313 467 35+	·45
·46	·365 155 90	·361 707 62	·358 357 27	·355 098 03	·351 923 82	·348 829 25+	·46
·47	·398 068 17	·395 413 72	·392 831 04	·390 315 05−	·387 861 26	·385 465 66	·47
·48	·431 660 95+	·429 858 18	·428 102 39	·426 390 22	·424 718 70	·423 085 19	·48
·49	·465 714 30	·464 802 84	·463 914 59	·463 047 90	·462 201 28	·461 373 41	·49
·50	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·500 000 00	·50

EXPERIMENTAL DISCUSSION OF THE (χ^2, P) TEST FOR GOODNESS OF FIT.

By KARL PEARSON.

(1) *Introductory.* In the *Philosophical Magazine* for 1900* I published, I think for the first time, a test which has since been much used for statistical purposes and has come to be spoken of as the (χ^2, P) test. The problem I had in mind was the following one: Samples of N are taken from a population classed in v categories, the chance of drawing an individual from the sth category being p_s; how will the cell contents $n_1, n_2 \ldots n_s \ldots n_v$ of the samples distribute themselves, and what is the probability, P, that samples may be drawn which deviate more from the mean parental values than a given sample?

With certain limitations the answer was shown to lie in the distribution of χ^2 by the curve

$$\left.\begin{aligned} y &= y_0 (\tfrac{1}{2}\chi^2)^{\frac{1}{2}(v-3)} e^{-\frac{1}{2}\chi^2} \quad [d(\tfrac{1}{2}\chi^2)], \\ \text{where} \qquad \chi^2 &= \mathop{S}_{s=1}^{s=v} \frac{(n_s - Np_s)^2}{Np_s} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}.$$

A short table was given in my original paper, and a longer one shortly afterwards computed by W. Palin Elderton†, by means of which it was possible to compute P from a known χ^2 and v (or n' as it was termed at that date). In dealing with the formula for χ^2 the p_s's might be those of a real parental population from which the sample had been actually drawn, or a hypothetical population from which we question whether it could reasonably be supposed to have been drawn.

Some time after this (1911‡) I published a second paper which discussed the problem, whether two observed samples of the same or different sizes, but with the same number v of categories, could reasonably be supposed to have been drawn from the same parent population. The problem is straightforward if the parent population be known. If it be not known, what values can be used to represent it? With some diffidence I suggested the values of the combined samples might be used to replace the parental population values. It was shown that in either of these cases we might use the (χ^2, P) test entering with v cells ($= n'$ of the (χ^2, P) Tables). The point to be emphasised here is: That when we use the known parent population, the two samples cannot be written as a contingency table, for the marginal total of v cells is *not* the sum of the contribution of each sample to a given cell, thus it is not a marginal

* Vol. L. pp. 157—175.

† *Biometrika*, Vol. I. pp. 155—163. See also: *Tables for Statisticians and Biometricians*, Part I, Table XII, 3rd Edn. 1931.

‡ *Biometrika*, Vol. VIII. pp. 250—254.

total in the contingency table sense. For this reason it is in my view a mistake to look upon every comparison of two samples or two series as a contingency table; it only becomes a contingency table in form when we, ignorant of the sampled or parent population, make the doubtful hypothesis that the latter population can be replaced by the sum of the two samples. This is so often overlooked in the text-book treatment of the χ^2 test for the co-origin of two samples that it is desirable to emphasise it.

But the biserial table formed from two samples differs seriously in another respect from a true contingency table. In the latter case, according to my envisaging of it, we start with a parent population and, drawing individuals in succession from it, table them according to two (or it may be more) characters. Thus in successive samples, say of size N, it is not only the contents of the $\kappa \times \lambda$ cells formed by κ categories of one character and λ of the other which vary from sample to sample, but also the marginal totals. In such a case there is only one degree of constraint on the contingency, arising from the size of the sample N.

In a paper published in 1916* I dealt with what I termed "partial contingency," namely, cases in which *linear* relations between the contents of certain numbers of cells existed from sample to sample. I proved in such cases that not only the n' by which we enter the (χ^2, P) table must be reduced by the number of such linear relations, but the observed χ^2 itself might also according to the nature of these constraints have to be reduced. To express the matter algebraically let n_{st} be the sample frequency in the sth row and tth column cell of the sample, and if p_{st} be the chance of drawing an individual from the s, t cell of the parent population, the expected value in the s, t cell of the sample will be $\bar{n}_{st} = Np_{st}$, and if

$$\left.\begin{aligned} \phi^2 &= \underset{s,t}{S} \frac{(n_{st} - \bar{n}_{st})^2}{N\bar{n}_{st}} = \frac{1}{N}\chi^2, \text{ say,} \\ \text{we have} \qquad N(1+\phi^2) &= N + \chi^2 = \frac{S(n_{st}^2)}{\bar{n}_{st}} \end{aligned}\right\} \qquad \text{(ii).}$$

Now the value of ϕ^2 or χ^2 will depend entirely on what value we give to $\bar{n}_{st}$. My idea was to give it such a value that ϕ^2 would lead (subject of course to the error of random sampling) to a measure of the association of the two characters or variates in the table. In order to achieve this I assumed the parent population to have no contingency or association between its variates. In this case $\bar{n}_{st} = Np_{s.}\,p_{.t} = \bar{n}_{s.}\,\bar{n}_{.t}/N$, where $p_{s.}$ and $p_{.t}$ are the respective chances of an individual being drawn out of the sth row, and an individual being drawn out of the tth column of the parental population, and $\bar{n}_{s.}$ and $\bar{n}_{.t}$ are the respective relative frequencies of row and column for samples of size N. It will be seen that $\bar{n}_{s.}$ and $\bar{n}_{.t}$ are not so far the sum of the sth row and tth column of any particular observed sample. We have

$$N(1+\phi^2) = N + \chi^2 = NS\left(\frac{n_{st}^2}{\bar{n}_{s.}\,\bar{n}_{.t}}\right) \qquad \text{(iii).}$$

Now if we do not know the parent population, we may follow one of two courses:

* *Biometrika*, Vol. XI. pp. 145—158 and pp. 159—190.

(*a*) Assume $\bar{n}_{s\cdot}$ and $\bar{n}_{\cdot t}$ are constants, values for the unknown parent population, and determine on the basis of their being constants the mean and variance, etc. of ϕ^2 and χ^2 in terms of the unknown algebraic quantities typified by $\bar{n}_{s\cdot}$ and $\bar{n}_{\cdot t}$. Finally in the formula so reached we are compelled to insert in want of better information the values of the $\bar{n}_{s\cdot}$, $\bar{n}_{\cdot t}$ actually reached in the observed Table.

(*b*) Assume that $\bar{n}_{s\cdot}$ and $\bar{n}_{\cdot t}$ are for each individual sample replaced by the sample values; we thus get a different definition of ϕ^2 and χ^2, and the mean and variances etc. will not be the same in cases (*a*) and (*b*). For in the latter case we have to allow for the variation of $n_{s\cdot}$ and $n_{\cdot t}$ in the formula

$$N(1+\phi^2) = N + \chi^2 = NS\left(\frac{n_{st}^2}{n_{s\cdot}\, n_{\cdot t}}\right) \quad \text{....................(iv).}$$

ϕ^2 has been discussed from the standpoints of both (*a*) and (*b*) in a series of previous papers in this *Journal**. What I want to emphasise is that if we start with (iv) as a definition of ϕ^2 and sample a parent population by taking individuals out one at a time and recording their characters, we obtain samples in which there is no fixing of either marginal total column. On the other hand, if we draw two independent samples, say, of boys and girls for eye colours, one from a population of boys, another from one of girls, we are dealing with a wholly different method of sampling. We can form a spurious contingency table out of these two rows with $2v$ cells, but theoretically we are limited to v cells, as I showed in the original treatment of the problem, and little appears to be gained by saying we have introduced v conditions of constraint.

(2) *Goodness of Fit.* I now turn to the main topic of this paper, the application of the χ^2 test to the problem of "goodness of fit." Here again divergence of opinion seems to be largely based on difference of aim and definition.

Suppose we take a random sample from a population, the whole of which we cannot observe or measure. The object of the anthropologist or craniologist is to ascertain how far, when making comparison with samples from other racial series, he may replace the not-fully measured parent population by his sample. In other words, we have two populations A and B, and we have samples a and b from them. We want to ascertain how far we can suppose A and b to be alike by a consideration of whether a representing A is alike to b. We are bound by the conditions of affairs to observe a sample of A; it stands for us as A, but which is not really A. How far does the fact that it is a sample only of A preclude us from ascertaining whether a sample b of B could with any probability have been obtained from A. This is the everyday problem of anthropologist, sociologist and most statisticians, but it is also the problem of "goodness of fit," and it is indeed the problem by which I originally reached the (χ^2, P) test, only the paragraph dealing with it reads obscurely and has been largely overlooked. Let us suppose we have a parent population of v categories

* *Biometrika*, Vol. v. pp. 192—203 (1906, with J. Blakeman); Vol. x. pp. 570—573 (1915); Vol. xi. pp. 215—230 (1916, with A. W. Young), and this corrected, Vol. xii. p. 260.

with probabilities of $p_1 \ldots p_s \ldots p_v$, that a sample of N has been drawn with frequencies $n_1 \ldots n_s \ldots n_v$, and that the moment coefficients about any point of this sample have been found $\mu_1', \mu_2' \ldots \mu_t'$, the variates corresponding to the frequencies being $x_1 \ldots x_s \ldots x_v$. Then

$$n_1 x_1^u + n_2 x_2^u + \ldots + n_v x_v^u = N\mu_u',$$

but it will *not* $= NM_u'$, where M_u' is the uth moment coefficient of the parent population. Every fresh sample will give a fresh series of moment coefficients, which will *not* equal those of the unobserved or unknown parent population. There is, thus, in this case no question of the limitation of the "degrees of freedom." When does such limitation occur? Only when we know the parent population, and therefore its moments, and fit various curves to that population by aid of t of its moments. We then have t-linear relations among the cell frequencies, and must reduce our "degrees of freedom" by that number. But surely this is not what we usually require? We do not know the parent population. We know a sample of it. To this sample we fit a curve and our problem is: How far may we use this curve to represent the unknown parent population? How far will further samples give corresponding χ^2 and χ'^2 when compared with the true parent population and with the sample from it?

We here reach two very important points:

(*a*) The distribution of the fitted curve to any sample gives a far lower χ^2 when compared with the parent population than the raw sample from which it was constructed.

(*b*) The distributions of χ^2 for the parent population and of χ'^2 for the fitted curve of the sample deduced from any fresh samples are such: (i) that the mean difference of χ^2 and χ'^2 is small and (ii) that the correlation of χ^2 and χ'^2 is extremely high. I attempted to give a proof of this in my original paper, a proof which has been considered obscure, but should have more or less indicated what my problem was. It was not "goodness of fit" of the curve deduced from the sample to the individual frequencies of that sample, but that of the χ'^2 distribution treated as an approximation to the χ^2 distribution that I had in view. In other words, I was considering and still want to consider how far we may replace the unknown parent population by the frequencies of the smooth curve deduced from a sample. For this purpose I have in the present paper selected a normal curve to represent the parent population with a standard deviation of 10. Luckily—for it saved me much labour—Dr Egon S. Pearson had somewhat over 1000 samples of 15 drawn from such a population by aid of Tippett's Random Sampling Numbers *, and I am very grateful to him for allowing me the use of them. He had computed the proportional frequencies for a central tenth of the standard deviation and for thirty such tenths on either side of the central group. Such a distribution is not an exact normal distribution, but it is very close to it; thus its standard deviation was 10·0283 without correction, and 9·9909 with Sheppard's correction, both close enough to the

* *Tracts for Computers*, No. XV. Cambridge University Press.

value 10 of the actual curve. As a matter of fact we may consider this distribution the parent one; there is no special merit in considering it an exact normal curve. Out of Dr Pearson's 15,000 odd samplings from the above parent population I took eight basic samples, none of which covered the same ground, they were independent samples from the parent population. These samples were of sizes 600, 300, 150, 105, 60, 45, 30 and 15. I term these the *basic samples*. The actual frequencies occurring in each sample I term the Raw Basic Samples. Each of these eight samples was fitted with a normal curve and the frequencies recomputed from this normal curve. These distributions I term the Smooth or Graduated Basic Samples. Finally, for a purpose to be explained later, I reduced all the frequencies of the Smooth or Graduated Basic Samples to a total of thirty. These may be referred to as the Graduated Basic Samples reduced. The parental population was reduced to a total thirty. I then proceeded to take 100 samples of thirty from the data. These were independent of each other and of the eight basic samples, i.e. all resulted from completely independent drawings. Of course had time and energy permitted, it would have been advisable to have had a large number of basic samples of each size and more than 100 samples to compare with them, but what has been done involved the computing of 900 χ^2's, and that means much labour*. The χ'^2's obtained from the smooth basic samples were then compared with the χ^2's obtained from the same series of 100 samples as against the parent population, and eight correlation tables were thus obtained. The close relationship between the χ'^2 from a smoothed basic sample and the χ^2 from the parent population became at once manifest, and there were very few cases in which one of the hundred samples would on the measure of its probability have been rejected or retained as a sample of the parent population when it would not in the same way have been rejected or retained by any one of the smoothed basic samples. In other words, the curve provided by the basic sample is a "good fit" to the parent population, and to judge by the present experience we are reasonably safe in replacing the unknown parent population by a graduated basic sample. Now the moment coefficients of the basic samples are not the same as those of the parent population, nor have they the same values for each basic sample. What has happened is this, that in calculating the distribution of the χ^2's of the 100 samples we have replaced the parent population frequencies by those of the smoothed basic sample, and the result has shown that we shall not make many or frequent errors of judgment in so doing. I have only been able to take eight samples of different sizes as an illustration. Had it been possible to take 50 or 100 samples of each size, we should no doubt have seen the advantage of the large over the small basic sample with respect to its goodness in representing the parent population. As a matter of fact in two basic samples one of 30 may be better than one of 300, though with a large number of samples those of 300 would certainly as a whole be better than those of 30 †.

* The bulk of the computing work was done for me by Dr L. T. Woo, but Mr Georg Hansmann undertook one series.

† The basic samples of 15 and 30 are for their size extremely favourable, i.e. give results much more accordant with those of the parent population than would be anticipated.

In obtaining the values of χ^2 it was needful to limit the number of categories, and ultimately 15 categories were selected each of $\frac{3}{10}$ths of the standard deviation, namely:

	Below −19·5	−19·5 / −16·5	−16·5 / −13·5	−13·5 / −10·5	−10·5 / −7·5	−7·5 / −4·5	−4·5 / −1·5	−1·5 / +1·5
Central Values	—	−18	−15	−12	−9	−6	−3	0

	+1·5 / +4·5	+4·5 / +7·5	+7·5 / +10·5	+10·5 / +13·5	+13·5 / +16·5	+16·5 / +19·5	Above +19·5
Central Values	+3	+6	+9	+12	+15	+18	—

It was desirable to have a considerable number of categories, but 15 categories was rather a large number for samples of 30, and it might have been better to increase the number of individuals in the test samples, as the small frequency in some of the categories would militate against the theoretical justification of replacing binomials by normal curves. As this would apply equally to all the basic samples and to the parent population, and we were dealing so to speak with relative values of χ^2 and χ'^2, I do not think it will affect the validity of our results, as it will influence χ^2 to much the same extent as χ'^2. I had also in mind another reason for choosing $\bar{n}_s$ to be small, which will appear later.

(3) *Experimental Details.* Table I gives the actual data for each basic sample in four columns. The first of these gives the raw basic frequencies (R.B.F.), the second gives the frequencies (G.B.F.) of the graduation—in this case a normal curve—replacing the original sample (R.B.F.), the third column gives the corresponding frequencies of the parent population (P.P.F.), and the fourth column the reduced basal frequencies (R.G.F.) for a sample of 30.

In calculating the χ^2 between the parental population and a basic sample, raw or smoothed by its curve, the full number of the sample was used, the relative frequencies of the parent population being modified to give the total of the sample. In treating the basic sample in its turn as a parent population for the 100 samples of 30, the reduced graduated values of the basic sample were of course employed. In the computing of χ^2 and the resulting P, where given, there is no question of any constraint beyond the size of the sample. Each basic sample has its own moment coefficients, and they are not the same as those of the parent population, nor of another basic sample of the same size.

Table II gives the mean and standard deviation of the raw data from which the smoothed frequencies were computed.

Basic Samples: Raw Frequencies (R.B.F.), *Graduated Frequencies* (G.B.F.), *Relative Parental Population Frequencies* (P.P.F.) *and Graduated Frequencies reduced to* 30 (R.G.F.).

Variate Range	Sample of 600				Sample of 300				Sample of 150				Sample of 105				Variate Range
	R.B.F.	G.B.F.	P.P.F.	R.G.F.	R.B.F.	G.B.F.	P.P.F.	R.G.F.	R.B.F.	G.B.F.	P.P.F.	R.G.F.	R.B.F.	G.B.F.	P.P.F.	R.G.F.	
Below −19·5	16	15·41	15·35	·77	3	5·21	7·68	·52	4	3·31	3·84	·66	3	1·83	2·69	·52	Below −19·5
−19·5—−16·5	11	13·97	14·33	·70	5	5·64	7·16	·56	4	3·14	3·58	·63	—	2·00	2·51	·57	−19·5—−16·5
−16·5—−13·5	22	22·62	23·42	1·13	11	9·88	11·71	·99	6	5·20	5·86	1·04	6	3·52	4·10	1·01	−16·5—−13·5
−13·5—−10·5	34	33·62	35·01	1·68	12	15·70	17·50	1·57	3	7·90	8·75	1·58	4	5·61	6·13	1·60	−13·5—−10·5
−10·5—− 7·5	51	45·89	47·86	2·29	34	22·61	23·93	2·26	12	10·99	11·97	2·20	7	8·08	8·38	2·31	−10·5—− 7·5
− 7·5—− 4·5	54	57·53	59·84	2·88	27	29·52	29·92	2·95	18	14·02	14·96	2·80	10	10·54	10·47	3·01	− 7·5—− 4·5
− 4·5—− 1·5	60	66·89	68·42	3·34	36	34·92	34·21	3·49	11	16·39	17·10	3·28	12	12·42	11·96	3·55	− 4·5—− 1·5
− 1·5—+ 1·5	71	69·18	71·54	3·46	36	37·45	35·78	3·75	20	17·57	17·88	3·51	18	13·25	12·52	3·79	− 1·5—+ 1·5
+ 1·5—+ 4·5	70	67·79	68·42	3·39	37	36·39	34·21	3·64	15	17·27	17·10	3·46	15	12·79	11·96	3·65	+ 1·5—+ 4·5
+ 4·5—+ 7·5	69	60·38	59·84	3·02	26	32·06	29·92	3·21	18	15·56	14·96	3·11	6	11·16	10·47	3·19	+ 4·5—+ 7·5
+ 7·5—+10·5	52	49·33	47·86	2·47	27	25·59	23·93	2·56	11	12·86	11·97	2·57	7	8·81	8·38	2·52	+ 7·5—+10·5
+10·5—+13·5	40	37·06	35·01	1·85	20	18·52	17·50	1·85	12	9·73	8·75	1·95	7	6·29	6·13	1·79	+10·5—+13·5
+13·5—+16·5	18	25·53	23·42	1·28	13	12·14	11·71	1·21	8	6·76	5·86	1·35	5	4·06	4·10	1·16	+13·5—+16·5
+16·5—+19·5	12	16·16	14·33	·81	3	7·22	7·16	·72	4	4·30	3·58	·86	2	2·37	2·51	·68	+16·5—+19·5
Above +19·5	20	18·64	15·35	·93	10	7·15	7·68	·72	4	5·00	3·84	1·00	3	2·27	2·69	·65	Above +19·5

Variate Range	Sample of 60				Sample of 45				Sample of 30				Sample of 15				Variate Range
	R.B.F.	G.B.F.	P.P.F.	R.G.F.	R.B.F.	G.B.F.	P.P.F.	R.G.F.	R.B.F.	G.B.F.	P.P.F.	R.G.F.	R.B.F.	G.B.F.	P.P.F.	R.G.F.	
Below −19·5	1	1·48	1·54	·74	3	1·91	1·15	1·27	1	·70	·77	·70	1	·25	·38	·50	Below −19·5
−19·5—−16·5	2	1·34	1·43	·67	2	1·39	1·07	·93	2	·71	·72	·71	1	·27	·36	·54	−19·5—−16·5
−16·5—−13·5	2	2·17	2·34	1·08	2	2·07	1·76	1·38	1	1·19	1·17	1·19	—	·48	·58	·96	−16·5—−13·5
−13·5—−10·5	—	3·23	3·50	1·62	2	2·84	2·63	1·89	1	1·82	1·75	1·82	—	·83	·88	1·66	−13·5—−10·5
−10·5—− 7·5	9	4·43	4·79	2·21	2	3·64	3·59	2·43	—	2·53	2·39	2·53	—	1·08	1·20	2·16	−10·5—− 7·5
− 7·5—− 4·5	5	5·59	5·98	2·80	5	4·31	4·49	2·88	2	3·18	2·99	3·18	1	1·49	1·50	2·98	− 7·5—− 4·5
− 4·5—− 1·5	4	6·50	6·84	3·25	4	4·76	5·13	3·17	7	3·63	3·42	3·63	2	1·78	1·71	3·56	− 4·5—− 1·5
− 1·5—+ 1·5	11	6·89	7·16	3·45	4	4·87	5·36	3·25	4	3·74	3·58	3·74	2	1·91	1·78	3·82	− 1·5—+ 1·5
+ 1·5—+ 4·5	5	6·77	6·84	3·38	3	4·64	5·13	3·09	2	3·50	3·42	3·50	1	1·85	1·71	3·70	+ 1·5—+ 4·5
+ 4·5—+ 7·5	8	6·11	5·98	3·06	5	4·10	4·49	2·73	5	2·98	2·99	2·98	3	1·62	1·50	3·24	+ 4·5—+ 7·5
+ 7·5—+10·5	3	5·07	4·79	2·53	6	3·36	3·59	2·24	2	2·29	2·39	2·29	3	1·28	1·20	2·56	+ 7·5—+10·5
+10·5—+13·5	1	3·86	3·50	1·93	3	2·56	2·63	·71	1	1·60	1·75	1·60	1	·91	·88	1·82	+10·5—+13·5
+13·5—+16·5	3	2·71	2·34	1·36	3	1·81	1·76	1·21	1	1·01	1·17	1·01	—	·59	·58	1·18	+13·5—+16·5
+16·5—+19·5	4	1·75	1·43	·87	—	1·19	1·07	·79	—	·58	·72	·58	—	·32	·36	·64	+16·5—+19·5
Above +19·5	2	2·10	1·54	1·05	1	1·55	1·15	1·03	1	·54	·77	·54	—	·34	·38	·68	Above +19·5

TABLE II.

Means and Standard Deviations of Basic Samples.

	Parent Population	Mean	Standard Deviation
		0	10*
Basic Samples	Size 600	+·4250	10·2275
	„ 300	+·6300	9·5291
	„ 150	+·9000	10·1380
	„ 105	+·4286	9·4334
	„ 60	+·8000	10·3252
	„ 45	−·5333	11·0041
	„ 30	−·5000	9·5795
	„ 15	+·6000	9·6867

It will be noted that only two of the means are negative; the odds against so small a number of negative signs are only about 5 to 1; yet should there be a series of rather improbable cases arising from Tippett's *Random Sampling Numbers*, we must remember that that series itself is a random sample, and may be a rather unusual one†. Table III compares the χ^2's and P's as found from the Raw Basic and Graduated Basic Samples as against the Parent Population. Two points at once

TABLE III.

Goodness of Fit of Basic Samples to Parent Population.

Basic Sample	Raw Basic Sample		Curve from Raw Basic Sample	
Size	χ^2	$P\ (n'=15)$	χ^2	$P\ (n'=15)$
600	8·2890	·8725	1·6789	·999,943
300	14·6609	·4024	2·2785	·999,682
150	9·8821	·7703	1·2936	·999,975
105	10·1715	·7491	·7782	>·999,999
60	19·6397	·1427	·5115	>·999,999
45	9·8996	·7691	1·0000	·999,999
30	12·3442	·5788	·1868	>·999,999
15	10·8809	·6951	·1462	>·999,999

* This was the standard deviation of the parental population curve from which the relative frequencies of this population were calculated. Working back from these computed frequencies to their standard deviation, we find 10·0288 for its value, = 9·9909 on applying Sheppard's correction. Corresponding to this the standard deviations recorded are all corrected values, and the relative frequencies of the basic samples were computed from the means and these corrected standard deviations.

† This caution is not given wholly unadvisedly. I have not myself made much use of Tippett's numbers, but recently I obtained in 100 trials *three* such unusual samples that only *one* should have occurred in 1,000,000 trials.

result from this table. First all the raw basic samples are, as of course they really must be, probable samples individually and as a group from the parent population. Secondly the curves fitted from the raw basic samples to the parent population—note, not to the raw basic frequencies themselves—are most excellent fits. They can be said to represent with a high degree of accuracy the parent population. The experience represented must, I think, be of interest and of real value to the anthropologist, who can rarely if ever measure whole populations, but has always before him the problem of whether a certain sample can be considered as belonging to a population he only knows from the graduated frequencies of another sample. We are not concerned here with the goodness of fit of a graduated curve to its raw sample, but of the goodness of fit of a graduated curve based on a raw sample to a graduated parent population from which the raw sample has been drawn. What we are considering in this case is the goodness of fit of a graduated sample to a graduated normal population, there is no limitation of the degrees of freedom, for the moments by which the graduation is determined change from sample to sample*.

(4) *Goodness of Fit of Graduated Basic Samples to Raw Basic Samples.* Here there is a point to which attention is not always given, or, perhaps, not sufficient attention. Many years ago† I showed that if two samples n_s, N, n_s', N'—s corresponding to the sth category out of v categories—were taken from a parent population in which p_s was the chance of drawing an individual at random from the sth category, then if

$$\chi^2 = \underset{s=1}{\overset{v}{S}} \frac{NN'}{N+N'} \frac{\left(\frac{n_s}{N} - \frac{n_s'}{N'}\right)^2}{p_s} \quad \text{.........................(v)},$$

we have χ^2 distributed according to the curve

$$y = y_0 e^{-\frac{1}{2}\chi^2} (\tfrac{1}{2}\chi^2)^{\frac{1}{2}(v-3)} \quad [d(\tfrac{1}{2}\chi^2)] \quad \text{....................(vi)}.$$

But an essential condition of this result is *that the series p_s is to be considered constant throughout the series of pairs of samples.* It is only under these conditions that the constants of χ^2, for example its mean and standard deviation, can be supposed given by the above distribution‡. In applying the χ^2 test to two samples, it is always well to consider what we are assuming our parent population to be. We may of course put for p_s any series of values we please, and can find the probability that the two samples belong to the corresponding population. If we have no knowledge of the parent population, we can use as the best substitute available for p_s the sum of the two samples, but our result is bound to be unreliable if those samples are not considerable.

* I may note here that I have often been asked: What is the value of so much curve fitting to samples? The answer is more or less conveyed in the present paper where we can see that the graduated basic sample effectively represents a parent population, even in the case of relatively small samples, and so serves as a standard for measuring the degree of divergence of one population from a second. This should be done not by comparing raw, but by comparing graduated basic samples.

† *Biometrika*, Vol. VIII. pp. 250—254, 1911.

‡ There is a still further limitation, n_s and n_s' must not be correlated.

On this latter assumption the value taken for χ^2 will be

$$\overset{v}{\underset{s=1}{S}} \frac{NN'\left(\frac{n_s}{N} - \frac{n_s'}{N'}\right)^2}{n_s + n_s'}$$

But if we do this we must remember that if we take another pair of samples indicated by $\tilde{n}_s$, $\tilde{n}_s'$, the corresponding

$$\chi^2 \text{ is not } \overset{v}{\underset{s=1}{S}} \frac{NN'\left(\frac{\tilde{n}_s}{N} - \frac{\tilde{n}_s'}{N'}\right)^2}{\tilde{n}_s + \tilde{n}_s'}$$

but is equal to
$$\overset{v}{\underset{s=1}{S}} \frac{NN'\left(\frac{\tilde{n}_s}{N} - \frac{\tilde{n}_s'}{N'}\right)^2}{n_s + n_s'},$$

otherwise the distribution of χ^2 is not given by

$$y = y_0 e^{-\frac{1}{2}\chi^2} (\tfrac{1}{2}\chi^2)^{\frac{1}{2}(v-3)} \quad [d(\tfrac{1}{2}\chi^2)].$$

Hence obscurity seems to me to arise when we write

n_1	n_2	...	n_s	...	n_v	N
n_1'	n_2'	...	n_s'	...	n_v'	N'
$(N+N')p_1$	$(N+N')p_2$	...	$(N+N')p_s$	...	$(N+N')p_v$	$N+1$

and replacing the horizontal totals by $n_s + n_s'$ speak of the result as a "contingency table" with v constraints. It is true that if we write

n_1	n_2	...	n_s	...	n_v	N
n_1'	n_2'	...	n_s'	...	n_v'	N'
$n_1 + n_1'$	$n_2 + n_2'$	...	$n_s + n_s'$	...	$n_v + n_v'$	$N+1$

this single pair of samples forms a contingency table* with $2v$ cells where the previous method has only v cells and we speak of $2v$ cells with v degrees of constraint. But the next or any other pair of samples will give

$\tilde{n}_1$	$\tilde{n}_2$	$\tilde{n}_s$	$\tilde{n}_v$	N
$\tilde{n}_1'$	$\tilde{n}_2'$			N'
$n_1 + n_1'$	$n_2 + n_2'$	$n_s + n_s'$	$n_v + n_v'$	$N + N'$

and although the horizontal totals are still fixed, this is far from being a contingency table unless $\tilde{n}_s + \tilde{n}_s' = n_s + n_s'$ for *every* pair of samples.

The relation $\chi^2 = (N + N')\phi^2$ holds for the first pair of samples, and this only when we replace $(N + N')p_s$ by $n_s + n_s'$. Such a relation as that written down leads the student to believe that χ^2 is always proportional to ϕ^2; for every other pair of samples beyond the first pair this is not true, and accordingly it seems to me a misfortune to discuss the matter under the heading of a contingency table; it confuses in the mind of a student the difference between a χ^2 test, with its pseudo-

* The equality of $(N+N')\phi^2$ and the χ^2 above is of course easily demonstrable.

contingency table based on a narrow hypothesis, and the true sample contingency table, where the marginal totals vary, and the only limitation is the single one, i.e. the size of the whole table. For these reasons I much prefer not to look upon the two sample test as a case of a contingency table, but as a comparison of the difference of the relative frequency of two samples with a certain parent population. Naturally this leads the student to define clearly what his parent population is supposed to be.

Now in Table III we have given two illustrations of "Goodness of Fit." First we have the raw basic samples and we compare them with the parent population for 15 categories. There is no doubt in this case that there is no limitation in the way of constraints beyond the size of the raw basic samples, and we look up Elderton's Table with $n' = 15$. Secondly we have tested the graduated against the parent population and we have used two moments, not of the parent population but of the raw sample; clearly except for the fractions such a sample could arise directly from sampling the parent population. Such a sample would be rare, and its goodness of fit is made obvious by the smallness of its χ^2. But we have no constraint; the moments of the next sample will differ from those of this one. We have by fitting by moments only selected one of the possible samples of the parent population, and we find that there are few samples better than it with regard to the fit to the parent population. To get—in particular from a small sample—the best possible approach to the parent population may be a difficult problem, but whether we graduate by two or four moments we are not restricting the number of degrees of freedom, we are simply selecting a possible sample out of endless possible samples. Where then does restriction of the degrees of freedom arise? Only as far as I can see when we fit by the moments of a sample a series of curves to the sample using the same moments in each case and the same number of categories; then the curve with the lowest χ^2 will have the best fit. But the P of the χ^2 table must be looked up under n' less the number of moments used*. This is however not the case I personally have had in view when considering "goodness of fit"; I want to ascertain how close the graduated sample is to the parent population, not to its raw sample. How far in the case of unknown parent populations can a measure of further samples from the parent population given by χ^2's be replaced by χ'^2's the measure of departure of these samples from the graduated first sample? This problem will be answered in the next section; the object of the present section is to consider the graduated basic sample in relation to the raw basic sample. In our case the graduated sample has been fitted by two moments: are we to reduce n' by two constraints? Clearly we are not fitting a series of curves to the one sample, we have selected a normal curve and we are not asking whether it is a better fit than a parabola or a sine curve. We must first determine which curve in the present comparison is to stand as the parent population. Obviously it cannot be the raw basic sample, for in that case it may have zero frequency in certain cells and

* For example, we might fit as graduation to a sample either the curve $y = y_0 e^{-\rho x} T_m(x/b)$ or the curve of Type IV $y = y_0 e^{-\gamma x}/(a^2 + x^2)^n$; then if the categories were n' in number we should have to reduce n' by four in ascertaining their P's.

accordingly the frequency of the normal curve could never have been obtained from it: that is, the normal curve would be an impossible sample. We must treat the graduated basic sample as our parent population, and ask what is the probability that the raw basic sample could be drawn from a parent population with the relative frequencies of the graduated sample. How we have obtained that parent population, whether by guess-work or by moments in any number, does not come into the problem. Here is a parent population, and here again is a sample which could be drawn from it: what is the probability P of samples like the present or more remote? It seems to me that this is a reasonable problem, and that it is the problem we usually desire to answer in curve fitting, rather than the question of the comparative fit of two curves determined by the same number of moments. If so, the process by which we have reached our parent population is a matter of indifference, we have no restriction of our degrees of freedom, beyond the size of the sample. I get my graduated basic sample—not to test it against other processes of graduation—but to see how far it may replace the unknown parent population from which the raw basic sample was drawn, and I do this by testing 100 experimental samples from a known parent population against that population and against the graduated basic sample to ascertain what is the relation of their χ^2's.

Table IV gives the Goodness of Fit of the Raw to the Graduated Basic Samples in the cases of the eight basic samples. It will be seen at once that the fit is good, and it should be, because our samples are owing to the equality of moments good ones—but the fit is nothing like as good as the fit of the graduated basic samples

TABLE IV.

Fit of Raw Basic to Graduated Basic Samples.

Size of Basic Sample	Raw Basic to Graduated Basic Samples			Raw Basic Samples to Parent Population		
	χ^2	P	Order	χ^2	P	Order
650	7·2899	·9216	1st	8·2890	·8725	1st
300	13·0665	·5214	7th	14·6609	·4024	7th
150	8·8009	·8427	3rd	9·8821	·7703	2nd
105	10·5628	·7193	4th	10·1715	·7491	4th
60	18·8529	·1711	8th	19·6397	·1427	8th
45	7·3810	·9174	2nd	9·8996	·7691	3rd
30	11·9425	·6109	5th	12·3442	·5788	6th
15	12·1290	·5960	6th	10·8809	·6951	5th

to the true parent population. This table shows results of considerable importance. The two orders are very nearly the same, the only interchanges being that of the 2nd and 3rd into the 3rd and 2nd, and of the 5th and 6th into the 6th and 5th. As we see the Basic Sample of 300 was a bad one and that of 45 an especially good one. No doubt had we been able to take a large number of basic samples

of 300 and of 45, these results would have been averaged out, and samples of 300 and of 45 put in more appropriate order. Table III shows us that for practical purposes there is very little to choose between the fits of all eight graduated basic samples to the parent population, we cannot place them in order without recalculating the (χ^2, P) table to more figures; all however give us a fit measured by $P > \cdot 9996$. The same four stand at the top in both orders and the same four at the bottom. As a *rough* rule we may therefore say that the raw sample which fits best its own graduation fits best the parent population. In other words, if we are seeking the "best" out of a number of samples from an unknown population, that best will be roughly indicated by the degree of goodness of fit it bears to its own graduation. Most investigators would say: "Oh, but a sample of 300 must give a better representation of an unknown parent population than one of 45!" It is of course true in the long run that we shall get better results from samples of 300 than from samples of 45. But in the present instance we have a case in which an individual sample of 45, both in its raw (Table IV) and graduated form (Table III), is a better fit to the parent population than a sample of 300. Of course it is needful for both samples to be true random samples from the parent population, not in any way selected for graduation and they must be graduated by the same process.

(5) *Parent Population and Graduated Basic Samples tested against* 100 *further Samples of* 30 *drawn from the Parent Population.* This is the main part of our experimental work, wherein we strive to determine the degree of accuracy with which the graduated basic sample can be used as representative of the parent population. It is in its turn to be treated as a parent population and the 100 samples from the original parent population will be tested by their χ^2 from the latter population and by their χ'^2 from the graduated basic sample as a spurious or step-parental population.

We shall investigate (*a*) the mean difference of $\chi^2 - \chi'^2$, (*b*) its standard deviation $\sigma_{\chi^2 - \chi'^2}$ and (*c*) the correlation of χ'^2 and χ^2, $r_{\chi'^2, \chi^2}$. As the range of χ^2 and χ'^2 is very considerable, and the correlation tables could only be formed and published for fairly considerable subranges, these were taken as unity for χ^2 and χ'^2. To determine the mean and standard deviation of $\chi^2 - \chi'^2$, the actual differences were taken and grouped in subranges of 0·2. Thus we find that mean $(\chi^2 - \chi'^2)$ is not exactly equal to mean χ^2 − mean χ'^2 as given by the correlation tables, nor

$$\sigma^2_{\chi^2 - \chi'^2} = \sigma^2_{\chi^2} + \sigma^2_{\chi'^2} - 2\sigma_{\chi}\sigma_{\chi'} r_{\chi^2, \chi'^2}$$

as given by the same correlation tables. The accordance however is good. Correlation Tables A—H tabulate the experience, and Table V gives the chief constants obtained in the manner described above.

The regressions are approximately linear, and accordingly the constant

$$\sigma_{\chi'^2}\sqrt{(1 - r^2_{\chi^2, \chi'^2})}$$

as well as the regression coefficient of χ'^2 on χ^2 or $R_{\chi^2, \chi'^2} = \sigma_{\chi'^2} r_{\chi^2, \chi'^2} / \sigma_{\chi^2}$ have been added.

TABLE V.

Constants of the Distributions of χ^2 for the Parent Populations and of χ'^2 for the Basic Samples as Step-Parent Populations.

Size of Basic Sample	From Distributions of $\chi^2-\chi'^2$ in 0·2 intervals		From the Correlation Tables A—H grouped for χ^2 and χ'^2 in unit intervals				
	Mean $\chi^2-\chi'^2$	$\sigma_{\chi^2-\chi'^2}$	Mean χ'^2	$\sigma_{\chi'^2}$	r_{χ^2,χ'^2}	$\sigma_{\chi'^2}\sqrt{1-r^2_{\chi^2,\chi'^2}}$	R_{χ^2,χ'^2}
600	+ ·018 (ii)	·7889 (i)	12·66	4·6402	·985,727 (i)	·7812	·988,646
300	− ·996 (vi)	1·6212 (v)	13·61	5·4474	·979,674 (ii)	1·0927	1·156,107
150	− ·398 (iv)	1·3046 (iii)	13·07	4·9135	·966,350 (iii)	1·2639	1·026,296
105	+1·938 (viii)	1·6578 (vi)	13·82	5·5222	·966,108 (iv)	1·4255	1·153,149
60	− ·164 (iii)	1·2521 (ii)	12·90	4·8017	·964,567 (v)	1·2669	1·001,094
45	+ ·017 (i)	1·6738 (vii)	12·71	4·2264	·934,983 (viii)	1·4991	·854,126
30	− ·429 (v)	1·5585 (iv)	13·15	5·1229	·939,327 (vii)	1·7573	1·040,112
15	− 1·280 (vii)	1·9747 (viii)	13·99	5·7277	·952,176 (vi)	1·7501	1·178,813

For the Parent Population: Mean $\chi^2 = 12{\cdot}68$, $\sigma_{\chi^2} = 4{\cdot}6265$.

Now examining this table, it will be seen that the average difference between χ'^2 and χ^2 is small, and that the variation in the difference is not great. The sixth column of the correlation coefficients shows how highly χ^2 and χ'^2 are correlated, the lowest correlation occurs with the sample of 45, but even this is greater than ·93.

TABLE VI.

Regression Lines of χ'^2 on χ^2.

Size of Sample	Regression Equation	Size of Sample	Regression Equation
600	$\chi'^2 = 0{\cdot}12 + {\cdot}98865\chi^2 \pm 0{\cdot}53$	60	$\chi'^2 = 0{\cdot}21 + 1{\cdot}00109\chi^2 \pm 0{\cdot}85$
300	$\chi'^2 = -1{\cdot}05 + 1{\cdot}15611\chi^2 \pm 0{\cdot}74$	45	$\chi'^2 = 1{\cdot}88 + {\cdot}85413\chi^2 \pm 1{\cdot}01$
150	$\chi'^2 = 0{\cdot}06 + 1{\cdot}02630\chi^2 \pm 0{\cdot}85$	30	$\chi'^2 = -0{\cdot}04 + 1{\cdot}04011\chi^2 \pm 1{\cdot}19$
105	$\chi'^2 = -0{\cdot}80 + 1{\cdot}15315\chi^2 \pm 0{\cdot}96$	15	$\chi'^2 = -0{\cdot}96 + 1{\cdot}17881\chi^2 \pm 1{\cdot}18$

In any case we can deduce χ'^2 from χ^2, or χ^2 from χ'^2 with greater accuracy than we can find in human beings any character of the right side from a knowledge of it on the left side; for example, the length of a right thigh bone from a knowledge of the length of the same bone on the left. I think any one who studies the correlation coefficients, the correlation tables and the regression lines will agree that as far as determining whether a sample B comes from an unknown population A, of which we have only a sample C, say of 60 or upwards, we shall rarely be wrong in our diagnosis, if we ask whether B could have been drawn from the sample C after graduation. That is to say, in the value of χ^2,

$$\chi^2 = \mathop{S}_{s=1}^{v} \frac{(n_s - \bar{n}_s)^2}{\bar{m}},$$

we replace the unknown $\bar{n}_s$'s series by $\tilde{n}_s$'s, where the latter are drawn from a graduated sample of the unknown population.

There is however a point to be noticed here. The distribution of χ^2 given by $y = y_0 e^{-\frac{1}{2}\chi^2}(\frac{1}{2}\chi^2)^{\frac{1}{2}(v-3)}$ depends on the $\bar{n}_s$'s being the means of the n_s's, and this is not correct, although a comparison between P.P.F. column under "Sample of 30," with the R.G.F. columns in all the samples in Table I will show that the differences are not large. Now the mean χ^2 in samples with $v = 15$ categories ought to be $v - 1 = 14$, and $\sigma_{\chi^2} = \sqrt{2(v-1)} = 5{\cdot}2915$. A consideration of columns 4 and 5 of Table V shows that the χ'^2's from Graduated Basic Samples only approach these values very roughly. The mean of their means for χ'^2 is 13·2389 and their mean $\sigma_{\chi'^2} = 5{\cdot}05025$. It might be plausibly argued that this is due to $\tilde{n}_s$ not being $\bar{n}_s$, but when we come to the sampling from the parent population where $\bar{n}_s$ has actually been used we find instead of a better correspondence a worse one, namely, mean $\chi^2 = 12{\cdot}68$ in place of 14, and $\sigma_{\chi^2} = 4{\cdot}6265$ in place of 5·2915*. It seems impossible therefore to attribute the divergence to $\tilde{n}_s$ not being equal to $\bar{n}_s$. There are two explanations which may account for the non-fulfilment by mean χ^2 and σ_{χ^2} of the theoretical values. First we have used 15 categories throughout, and this seemed a necessity for purposes of comparison; further, the categories were not unsuitable when we were comparing the larger samples directly with the parent population or with one another in their raw and graduated forms. But it is not so satisfactory when we test the parent population or the graduated samples against the samples of 30, as the theoretical values in some of the categories become small. Some experimental work, however, seems to indicate that not very great effect is produced by the small categories. The second point is that it is due to the approximate nature of the curve $y = y_0 e^{-\frac{1}{2}\chi^2}(\frac{1}{2}\chi^2)^{\frac{1}{2}(v-3)}$.

While the true $\bar{\chi}^2 = v - 1$, the deduction of the variance of χ^2 as $2(v-1)$ depends on the above curve being applicable, which actually it is not. There is a limit to the value of χ^2 which is, I think, $\chi_l^2 = \dfrac{N(N - \bar{n}_l)}{\bar{n}_l}$, where N is the size of the sample, and $\bar{n}_l$ the *least* relative cell frequency of the parental population†. Thus in order to get the customary equation for χ^2 with an infinite range we require to make $\bar{n}_l$ as small as possible, but to do so is to disregard the principle that $\bar{n}_s$ must be relatively large compared to N in order that we may replace the binomial by a normal curve. We are thus thrust on the horns of a dilemma. If we say that $(\frac{1}{10} + \frac{9}{10})^N$ is the most skew binomial that can reasonably be represented by a normal curve, and we take $n_l = \frac{1}{10}N$, then $\chi_l^2 = 9N$, and for a small sample we may doubt whether it is legitimate to treat this as an infinite range.

* The divergences are of course not impossible, but they point in one way; actually they are for difference of means 1·32 ± ·36 and for difference of standard deviations ·665 ± ·252 approximately.

† In the case of the 100 samples of 30 compared with the parent population, $\chi_l^2 = 1220$, and this range might be treated as approximately infinite for our purposes, but to obtain it we have infringed the condition as to replacing binomials by normal curves.

The actual value of the variance of χ^2, when in Equation (i) the $\bar{n}_s = Np_s$'s are the true mean of the n_s's, is given by*

$$\sigma^2_{\chi^2} = 2(v-1)\left(1-\frac{1}{N}\right) - \frac{v^2}{N} + S\left(\frac{1}{\bar{n}_s}\right) \quad \text{.................(vii).}$$

Hence the usual value

$$\sigma^2_{\chi^2} = 2(v-1)$$

may be modified in two ways: (*a*) if the sample be small, and v be not small, the negative term v^2/N may be by no means negligible; for example, if $v=15$ and $N=30$, the term $v^2/N = 7{\cdot}5$, which cannot be neglected as compared with 27·07; (*b*) on the other hand the term $S\left(\frac{1}{\bar{n}_s}\right)$ is additive, and if we are dealing with a small sample and with a fair sized v, this may be considerable. For example, in the case of our experimental parent population reduced to a size of 30, $S\left(\frac{1}{\bar{n}_s}\right) = 10{\cdot}597$, so that the theoretical variance in that case is 30·17 instead of 28, giving $\sigma_{\chi^2} = 5{\cdot}493$ instead of 5·292, and differing still more from the observed 4·627, the deviation being 3·4 times its probable error. Even with a sample of 50 in 10 categories so chosen that no category contains less than 4, and thus $S\left(\frac{1}{\bar{n}_s}\right)$ reduced to a minimum, the term v^2/N will still be 2, and this is not negligible as compared with 15·68. The use of $\sigma^2_{\chi^2} = 2(v-1)$ and consequently of (vi) in the case of *small* samples is certainly to be deprecated.

I again hazard the suggestion that the better distribution of χ^2 in such cases is to be found from the curve

$$y = y_0 (\tfrac{1}{2}\chi^2)^{p_1} (\tfrac{1}{2}\chi_l^2 - \tfrac{1}{2}\chi^2)^{p_2} \quad \text{........................(viii),}$$

where

$$\chi_l^2 = N(N-\bar{n}_l)/\bar{n}_l,$$

$$p_1 + 1 = \frac{\bar{\chi}^2}{\chi_l^2}\left\{\frac{\bar{\chi}^2(\chi_l^2 - \bar{\chi}^2)}{\sigma^2_{\chi^2}} - 1\right\},$$

$$p_2 + 1 = \left(1 - \frac{\bar{\chi}^2}{\chi_l^2}\right)\left\{\frac{\bar{\chi}^2(\chi_l^2 - \bar{\chi}^2)}{\sigma^2_{\chi^2}} - 1\right\},$$

and

$$\bar{\chi}^2 = (v-1), \qquad \sigma^2_{\chi^2} = 2(v-1)\left(1-\frac{1}{N}\right) - \frac{v^2}{N} + S\left(\frac{1}{\bar{n}_s}\right).$$

The *Table of the Incomplete* B-*Function* will provide the requisite probability P for a given χ^2.

We can see easily how (viii) passes into a Type III curve if χ_l^2 be large. In that case we have approximately

$$\left.\begin{aligned} p_1 &= \frac{(\bar{\chi}^2)^2}{\sigma^2_{\chi^2}} - 1 = \tfrac{1}{2}\frac{v-1}{\lambda} - 1 \\ p_2 &= \frac{\bar{\chi}^2\chi_l^2}{\sigma^2_{\chi^2}} = \frac{\chi_l^2}{\lambda} \end{aligned}\right\} \quad \text{........................(ix),}$$

* *Biometrika*, Vol. XI, Equation (xii), for the value from a limited parent population, and $\bar{n}_s$ not necessarily the mean of n_s, and as above for the case of $\bar{n}_s$ = mean of n_s.

where
$$\lambda = 1 - \frac{1}{N} - \frac{v^2}{2(v-1)N} + \frac{1}{2(v-1)} S\left(\frac{1}{\bar{n}_s}\right).$$

Hence
$$y = y_0' \left(\tfrac{1}{2}\chi^2\right)^{\frac{1}{2}\frac{v-1}{\lambda} - 1} \left(1 - \frac{\frac{1}{2}\chi^2}{\frac{1}{2}\bar{\chi}_l^2}\right)^{\frac{1}{2}\bar{\chi}_l^2 \frac{1}{\lambda}} \quad \text{..................(x)},$$
or since χ_l^2 is large we reach
$$\left.\begin{aligned} y &= y_0' \left(\tfrac{1}{2}\chi^2\right)^{\frac{1}{2}\frac{v-1}{\lambda} - 1} e^{-\frac{1}{2}\frac{\chi^2}{\lambda}} \\ &= y_0'' \left(\tfrac{1}{2}\frac{\chi^2}{\lambda}\right)^{\frac{1}{2}\frac{v-1}{\lambda} - 1} e^{-\frac{1}{2}\frac{\chi^2}{\lambda}} \end{aligned}\right\} \quad \text{.....................(xi)}.$$

It is accordingly $\frac{1}{2}\frac{\chi^2}{\lambda}$ which is approximately given by a Type III curve, and the power is not $\frac{1}{2}(v-3)$, but $\frac{1}{2}\frac{v-1}{\lambda} - 1$; the probability P will be easily calculated from the *Incomplete Γ-Function Table*. In the case of the samples from the parent population of this paper $\lambda = 1{\cdot}07737$, and since $\chi_l^2 = 1220$, the transition to (xi) is reasonably legitimate. But the usual χ^2 will be in error to about 8 °/$_\circ$.

The justification for (viii) lies in the fact that it gives the true start and range of the χ^2 curve as well as its true mean and variance. It will probably account with considerable accuracy for the binomials not closely following a normal distribution; and with the Tables of the Incomplete Γ- and B-Functions the P corresponding to (viii) or (xi) may be obtained as quickly as from Palin Elderton's Table.

The whole subject is worthy of further experimental investigation, for if my conjecture as to the approximate accuracy of (viii) and (xi) be verified, the use of the χ^2 test could be extended to small samples and small cell frequencies, which are not suitable in the case of the ordinary (χ^2, P) process.

The fundamental experiment of the present paper is in part intended to illustrate the need for widening the nature of Equation (i). No discriminating investigation could be based on the present data without increasing much beyond 100 the number of samples taken.

(6) *Actual Comparison of the P's from a Parent Population and from a Graduated Sample from that Population.* My original intention was to publish side by side the P's determined from the χ^2's and χ'^2's of the Parent Population and the eight Graduated Basic Samples. But the large amount of labour and of printing involved in computing and publishing 900 P's induced me to confine my attention to a single sample, which I have taken a good way down the list to indicate that a relatively small basic sample, say 50 to 100, if graduated, will provide a reasonable indication of whether further samples do or do not belong to the unknown parent population. The 100 values of P for the 105 Basic Sample are given in Table VII.

The problem turns here on how many samples which belong really to the parent population would have been rejected by the graduated basic sample. Suppose first we take the 2 °/$_\circ$ standard. No. 82 would be rejected by both P and P' tests if

TABLE VII.

Comparison of the Probabilities of 100 *Samples drawn from a Parent Population and again supposed to be drawn from a Graduated Basic Sample of* 105.

Sample Index No.	*P* as drawn from Parent Population	*P'* as drawn from Basic Sample	Sample Index No.	*P* as drawn from Parent Population	*P'* as drawn from Basic Sample	Sample Index No.	*P* as drawn from Parent Population	*P'* as drawn from Basic Sample	Sample Index No.	*P* as drawn from Parent Population	*P'* as drawn from Basic Sample
1	·763	·754	26	·720	·710	51	·815	·804	76	·827	·710
2	·204	·125	27	·609	·581	52	·679	·637	77	·303	·200
3	·681	·584	28	·462	·434	53	·704	·718	78	·209	·046
4	·232	·153	29	·763	·700	54	·990	·971	79	·697	·484
5	·202	·281	30	·775	·820	55	·579	·376	80	·846	·872
6	·421	·221	31	·819	·718	56	·098	·024	81	·630	·546
7	·846	·872	32	·441	·453	57	·832	·746	82	·019	·007
8	·942	·932	33	·054	·066	58	·439	·526	83	·469	·270
9	·245	·165	34	·310	·283	59	·932	·962	84	·445	·229
10	·135	·057	35	·460	·360	60	·128	·041	85	·712	·553
11	·688	·725	36	·567	·409	61	·382	·373	86	·973	·982
12	·918	·879	37	·320	·307	62	·117	·026	87	·774	·830
13	·900	·746	38	·194	·240	63	·381	·461	88	·822	·766
14	·060	·071	39	·651	·615	64	·105	·019	89	·788	·721
15	·779	·657	40	·192	·087	65	·879	·873	90	·776	·728
16	·309	·169	41	·848	·780	66	·456	·305	91	·708	·499
17	·747	·625	42	·279	·161	67	·889	·937	92	·902	·873
18	·987	·976	43	·999	·999	68	·866	·835	93	·359	·233
19	·390	·414	44	·285	·404	69	·869	·894	94	·803	·644
20	·857	·854	45	·297	·338	70	·743	·721	95	·174	·072
21	·567	·598	46	·575	·492	71	·437	·383	96	·711	·722
22	·286	·167	47	·106	·060	72	·180	·112	97	·284	·241
23	·850	·895	48	·586	·678	73	·852	·806	98	·477	·310
24	·829	·791	49	·945	·949	74	·622	·424	99	·773	·778
25	·666	·652	50	·859	·884	75	·231	·163	100	·763	·754

an isolated sample. No. 64 would be retained as a sample of the parent population and rejected as a sample from the basic sample population, had it occurred as an isolated sample. Actually it or something worse might be expected to occur twice in 100 samples. Thus dealing with a 2 °/$_0$ limit and an isolated sample we should have made an error once in a hundred times in rejecting a sample from the parent population and twice in 100 times if we used the basic graduated sample in place of the parent population. If we used a 5 °/$_0$ level, No. 82 would be the only sample rejected on account of its *P* value in the case of the parent population, while Nos. 56, 60, 64, 78 and 82 would all be rejected on the basic sample test. The reader might hastily pass to the conclusion that the basic sample does not effectively represent the parent population. But the conclusion is rather that the present sampling is *too* favourable. A 5 °/$_0$ level means that there are *five* cases in the 100 below it, the parent population shows only *one*, while the graduated basic sample actually records five.

We may consider the matter from another standpoint. The distribution of the probability integrals of any continuous curve is a rectangle, every probability

between 0 and 1 being equally likely*. Accordingly the distribution of P and P' should be linear. Dividing into 10 groups we have the following scheme:

TABLE VIII.

Distribution of P and P'.

Probability =	$\frac{\cdot 00}{\cdot 10}$	$\frac{\cdot 10}{\cdot 20}$	$\frac{\cdot 20}{\cdot 30}$	$\frac{\cdot 30}{\cdot 40}$	$\frac{\cdot 40}{\cdot 50}$	$\frac{\cdot 50}{\cdot 60}$	$\frac{\cdot 60}{\cdot 70}$	$\frac{\cdot 70}{\cdot 80}$	$\frac{\cdot 80}{\cdot 90}$	$\frac{\cdot 90}{1\cdot 00}$	Total
Expected	10	10	10	10	10	10	10	10	10	10	100
Parent Population, P ...	5	9	11	8	10	5	9	17	16·5	9·5	100
Graduated Basic Sample, P'	9	10·5	9·5	6	11	6	8·5	16·5	16	7	100

In both cases we find a redundancy of rather favourable samples.

For the Parent Population: $\chi^2 = 15{\cdot}65$, $P = {\cdot}075$.

For the Basic Sample of 105: $\chi'^2 = 12{\cdot}40$, $P' = {\cdot}193$.

The odds in the first case are about 12 or 13 to 1, and in the second case about 4 or 5 to 1. Thus the Graduated Basic Sample gives the more reasonable result. Both are possible in the single isolated trial.

The reason for there being less correspondence between the P and P' for the series of 100 samples lies in the low standard deviation of the 105 Basic Sample; see Table II. It is the lowest of all eight samples. Hence in the case of rare individuals being drawn from the Parent Population, they would be still rarer in the case of the 105 Basic Sample, and accordingly what is a rare sample from the standpoint of the Parent Population will be still rarer in the case of the Basic Sample, i.e. when P is small, P' will be still smaller†.

The reader may ask for some evidence that the Normal Parent Population and the Basic Sample would correspond in the same manner when the samples tested were drawn not from the former but from an entirely different population. For this purpose I took a Rectangular Population, and not to protract matters too severely took only ten samples of thirty from it. The values of P for three assumed parent populations are given in Table IX. It will be seen at once that the values of P for both the Normal Parent Population and the Graduated Basic Sample of 105 as parent population are *on the whole* strongly against the samples from the Rectangular Population being their offspring; the Graduated Basic Sample is even more strenuous than the grand-parental normal curve population, owing to the fact of its smaller standard deviation. Naturally the bulk of the contributions to χ^2 come

* Thus Bayes' theorem applies accurately to such distributions of P, all chances being equally likely.

† This reduction of the standard deviation will be somewhat modified by the shifting of the mean, which is, however, nearly the smallest shift of the series, and this shift, if it compensates for the reduced standard deviation effect at one tail, will emphasise it at the other.

TABLE IX.

Comparison of values of P from Samples of 30 drawn from a Rectangular Population, with their Pure Population, the Normal Curve Parent Population, and the Graduated Basic Sample of 105.

Chance of occurrence if drawn from:	Index Number of Sample from a Rectangular Population								
	I	II	III	IV	V	VI	VII	VIII	IX
Rectangular Population as Parent	·6063	·8311	·5265	·0193	·0458	·2562	·6063	·4497	·3134
Normal Curve as Parent Population	·7418	·0054	·0060	$<$·000,0005	$<$·000,0005	$<$·000,004	·0239	·0024	·0011
Normal Curve of Basic Sample of 105 as Parent Population	·5607	·0010	·0045	$<$·000,0005	$<$·000,0005	$<$·000,005	·0072	·000,003	·000,170

from random variation in the extreme categories of the rectangle*. We may I think conclude that the Parent Population and the Graduated Basic Sample will give the same sort of judgment with regard to a population differing from either of them.

But having said this, we examine our table further and begin to realise the weakness of small samples. Two true samples (IV and V) out of ten from the rectangle would be rejected on the $P = {}$·05 basis, and one (IV) on the ·02 basis. On the other hand two samples out of ten would be accepted as genuine samples of the Normal Parent Population, and one out of ten as a genuine sample of the Basic Sample of 105 on the ·05 basis, and two from both on the ·02 basis. Indeed, Sample I is a better sample from a normal population than from a rectangular population, its true parent.

Of course such anomalies will occur, but if they can occur in ten samples from such very different distributions as a normal curve and a rectangle, must we not be somewhat anxious whether they will not occur, and more frequently occur, when we compare two small samples and assert identity of origin? The samples may really have arisen from two wholly different populations, but far more accordant than a rectangle and a normal curve!

To illustrate this point I will deduce, by aid of the formula†

$$\chi^2 = SNN' \frac{\left(\frac{f_s}{N} - \frac{f_s'}{N'}\right)^2}{f_s + f_s'} = S\frac{(f_s - f_s')^2}{f_s + f_s'}, \text{ if } N \doteq N',$$

* The sampling was done by putting into a box 10 of each of the letters $A, B, C \ldots M, N, O$ on tickets. A ticket was then drawn, its letter recorded, and it was returned to the box. The box had then its lid closed, it was waved about, rotated and shaken, and then a second ticket drawn, and the process continually repeated until 300 tickets had been recorded. The two exceptional samples (IV) and (V) arose from the last letter, O, occurring seven times, and the last but one, N, seven times. Observation showed that it was not through faulty shaking, as it was not the same ticket repeating itself. Further, it could hardly be due to "clustering", as the tickets had been introduced into the box in a manner which avoided this, and tickets of the same letter did not follow in approximate succession.

† *Biometrika*, Vol. VIII. p. 252.

the probability of our 7 raw basic samples of sizes 15, 30, 45, 60, 105, 150 and 300 actually taken from a normal population being drawn from a rectangular population.

Using the ungraduated raw samples, we have:

TABLE X.

Probability of Raw Samples from a Normal Population having a Rectangular Parent.

Size of Samples	15	30	45	60	105	150	300
Values of P	[illegible]3	·6783	·8306	·3244	·0326	·0051	$<$·000,0005

It will be clear from this table that even with parent populations so different as those here dealt with, the χ^2 test is inadequate to discriminate between raw samples from these populations, if the samples have not sizes of the order 100, and even then not on the ·02 probability basis. Safety may be said to begin between 105 and 150, and if 50 or below be said to be "small" sample sizes, it is not dogmatic to assert that the χ^2* test ought never to be applied to such small samples.

(7) *Conclusions.* An endeavour has been made in this paper to mark more clearly the distinctions the writer had in view in introducing χ^2 and ϕ^2 into statistical theory and practice.

(i) If χ^2 be defined as
$$\chi^2 = \underset{s=1}{\overset{s=v}{S}} \frac{(n_s - \bar{n}_s)^2}{\bar{n}_s}$$
then $\bar{n}_s$ is in a succession of samples a constant and equal to the mean value of n_s. If this condition be satisfied, χ^2 is given approximately, but only approximately, by the curve
$$y = y_0 e^{-\frac{1}{2}\chi^2} (\tfrac{1}{2}\chi^2)^{\frac{1}{2}(v-3)} \ [d(\tfrac{1}{2}\chi^2)],$$
where v is the number of cells $= n'$ of Elderton's (χ^2, P) table.

The mean of χ^2 is $v-1$, but its true standard deviation is not $\sqrt{2(v-1)}$ but is given by Equation (vii) above. It is suggested that either Equation (viii) or Equation (xi) will give a better value for the P corresponding to a given χ^2, using either the Incomplete B- or Γ-Function Tables, than Elderton's (χ^2, P) Table, when N is not very large or any $1/n_s$ is not negligible as compared to unity.

(ii) It has been pointed out that the main use of χ^2 was intended to be the comparison of a considerable graduated sample of a parent population with further samples in order to test whether such samples were or were not likely to be samples from the parent population, only known through this graduated sample. It is shown from a series of experimental examples that the χ'^2's and P''s from the graduated sample are very highly correlated with the χ^2's and P's from the parent

* I mean the form of the χ^2 test based on the distribution $y = y_0 e^{-\frac{1}{2}\chi^2} (\frac{1}{2}\chi^2)^{\frac{1}{2}(v-3)}$.

population, so that without frequent wrong judgment we may use the step-parental population in place of the unknown parent population. This amounts to using for $\bar{n}_s$ the values found from the graduated sample population They still remain constant throughout the comparison with further samples, and the larger the size of the graduated sample the higher on the average will be the correlation between the χ^2's from the step-parental and true parental populations.

(iii) Incidentally it is pointed out that even for small samples an immensely better P is obtained from a graduated than from a raw sample, and this even when the size of the sample is small.

(iv) More than twenty years ago I gave a test for two samples of sizes N and N', categories n_s, n_s' being drawn from the same parent population with relative frequency of the sth category p_s. It consisted in calculating

$$\chi^2 = \overset{v}{\underset{s=1}{S}} \frac{NN'}{N+N'} \frac{\left(\frac{n_s}{N} - \frac{n_s'}{N'}\right)^2}{p_s},$$

there being v categories in both samples, and then applying the (χ^2, P) table. Here p_s is supposed throughout the further sampling to be a constant.

If the parent population be supposed unknown, I suggested that the best value available was $p_s = (n_s + n_s')/(N + N')$; this would not be very accurate for small samples.

In this case
$$\chi^2 = \overset{v}{\underset{s=1}{S}} \frac{NN'\left(\frac{n_s}{N} - \frac{n_s'}{N'}\right)^2}{n_s + n_s'}.$$

But it has been frequently overlooked that in adopting this value of χ^2, we must in measuring P remember that in further samples the denominator $n_s + n_s'$ is supposed to remain constant, while we vary n_s and n_s' in the numerator, otherwise the distribution of χ^2 on which the (χ^2, P) table is based is incorrect. The whole matter has in my opinion been rendered unnecessarily obscure by writing the two samples in the form of a spurious biserial contingency table. This is said to have $2v$ cells, and to lack $v+1$ "degrees of freedom." If the first pair of samples gives a contingency table the second will not, and from this manner of approach we lose sight of the true difference between a real and a spurious contingency table. In the former all the constituents of the marginal totals are free, and the only limitation is the total size of the sample in the table.

(v) If we, however, write χ^2 in the form of

$$(N+N')\,\phi^2 = \overset{v}{\underset{s=1}{S}} \frac{\left(n_s - \frac{N}{N+N'}(n_s + n_s')\right)^2}{\frac{N}{N+N'}(n_s + n_s')} + \overset{v}{\underset{s=1}{S}} \frac{\left(n_s' - \frac{N'}{N+N'}(n_s + n_s')\right)^2}{\frac{N}{N+N'}(n_s + n_s')}$$

we are not only giving ourselves double work, but are apt to forget that to get the approximate equation for χ^2 we must consider $n_s + n_s'$ constant in the numerator,

which it is not if we take another pair of samples. There is nothing whatever to prevent our choosing

$$\phi^2 = \overset{v}{\underset{s=1}{S}} \frac{NN'}{N+N'} \frac{\left(\frac{n_s}{N} - \frac{n_s'}{N'}\right)^2}{n_s + n_s'}$$

as our measure of accordance of the two samples, making n_s, n_s' *vary in both numerator and denominator*. But if this be done, the distribution of ϕ^2 is not that of $\chi^2/(N+N')$, and it cannot be deduced from the (χ^2, P) table. Even approximate values of the mean and variance are complicated, and the experimental study of the distribution of ϕ^2 has only been started by Professor Kondo's recent paper.

(vi) If we take a sample, graduate it by aid of t moments and then compare it with any population, it is, apart from fractions of a unit, a possible sample from that population, and we are at liberty to look out P in the ordinary (χ^2, P) table and judge where it stands among other possible samples of the same size and the same categories. We have not limited anything by obtaining our graduated sample by moments. When we do limit by moments is when we fit a series of curves to a given distribution by moments, the curve moments being in each case those of the given distribution. In such a case we compare the relative goodness of fit of various curves obtained by t moments and our degrees of freedom are reduced by t. An application of such limitation of degrees of freedom was made by me in 1915 and applied to the case of death-rates. It was shown in the memoirs concerned with this topic that in certain cases not only the degrees of freedom, but the value of χ^2 with which the table was entered might need modification*. Another case of the modification of both v and χ^2 to get a better measure of P is indicated in this paper: see pp. 366—367 above.

(vii) We have also seen in this paper that it is still probably legitimate to calculate P from χ^2 when, owing to the smallness of the sample and of its categories, the distribution $y = y_0 e^{-\frac{1}{2}\chi^2} (\frac{1}{2}\chi^2)^{\frac{1}{2}(v-3)}$ is no longer accurate. In such cases we are thrown back on frequency curves which are generalisations of the usual χ^2-curve, and which can be integrated by aid of the *Incomplete* Γ-*Function* and the *Incomplete* B-*Function Tables*. There is here a field for much experimental work of a useful kind.

* *Biometrika*, Vol. XI, pp. 145—184.

TABLE A.

χ'^2 from Basic Sample of 600 and χ^2 from Parent Population.

Parent Population χ^2

Basic Sample χ'^2	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12	12—13	13—14	14—15	15—16	16—17	17—18	18—19	19—20	20—21	21—22	22—23	23—24	24—25	25—26	26—27	27—28	Totals
2—3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3—4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
4—5	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
5—6	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
6—7	—	—	—	—	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
7—8	—	—	—	—	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
8—9	—	—	—	—	—	—	7	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
9—10	—	—	—	—	—	—	4	6	4	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15
10—11	—	—	—	—	—	—	—	6	3	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	13
11—12	—	—	—	—	—	—	—	—	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
12—13	—	—	—	—	—	—	—	—	—	1	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
13—14	—	—	—	—	—	—	—	—	—	1	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
14—15	—	—	—	—	—	—	—	—	—	—	—	1	4	—	—	—	—	—	—	—	—	—	—	—	—	—	5
15—16	—	—	—	—	—	—	—	—	—	—	—	—	4	1	2	2	—	—	—	—	—	—	—	—	—	—	9
16—17	—	—	—	—	—	—	—	—	—	—	—	—	—	1	2	1	1	—	—	—	—	—	—	—	—	—	5
17—18	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4	—	1	—	—	—	—	—	—	—	—	—	5
18—19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	3	1	—	—	—	—	—	—	—	—	5
19—20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
20—21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	4	1	—	—	—	—	—	—	6
21—22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
22—23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
23—24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
24—25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
25—26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
26—27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
27—28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
28—29	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	—	1	2	1	2	4	12	15	9	7	5	5	8	2	8	4	6	1	4	1	—	2	—	—	—	1	100

TABLE B.

χ'^2 from Basic Sample of 300 and χ^2 from Parent Population.

Parent Population χ^2 (columns); Basic Sample χ'^2 (rows)

	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12	12—13	13—14	14—15	15—16	16—17	17—18	18—19	19—20	20—21	21—22	22—23	23—24	24—25	25—26	26—27	27—28	Totals
2—3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3—4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
4—5	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
5—6	—	—	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
6—7	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
7—8	—	—	—	—	—	1	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
8—9	—	—	—	—	—	1	6	5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12
9—10	—	—	—	—	—	—	1	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
10—11	—	—	—	—	—	1	2	5	5	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15
11—12	—	—	—	—	—	—	—	2	—	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
12—13	—	—	—	—	—	—	—	—	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
13—14	—	—	—	—	—	—	—	—	2	1	2	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	7
14—15	—	—	—	—	—	—	—	—	—	1	1	2	2	—	1	—	—	—	—	—	—	—	—	—	—	—	7
15—16	—	—	—	—	—	—	—	—	—	—	—	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	4
16—17	—	—	—	—	—	—	—	—	—	—	—	1	—	1	1	—	1	—	—	—	—	—	—	—	—	—	4
17—18	—	—	—	—	—	—	—	—	—	—	—	—	1	1	3	1	1	—	—	—	—	—	—	—	—	—	7
18—19	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	2	—	—	—	—	—	—	—	—	—	—	4
19—20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	2
20—21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	2
21—22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	1	—	—	—	—	3
22—23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	1	—	—	—	—	3
23—24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
24—25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	2
25—26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
26—27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	—	—	—	—	—	3
27—28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
28—29	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
29—30	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
30—31	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
31—32	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	—	1	2	1	2	4	12	15	9	7	5	5	8	2	8	4	6	1	4	1	—	2	—	—	—	1	100

TABLE C.

χ'^2 *from Basic Sample of* 150 *and* χ^2 *from Parent Population.*

Parent Population χ^2

Basic Sample χ^2	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12	12—13	13—14	14—15	15—16	16—17	17—18	18—19	19—20	20—21	21—22	22—23	23—24	24—25	25—26	26—27	27—28	Totals
2—3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3—4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
4—5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
5—6	—	—	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
6—7	—	—	—	—	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
7—8	—	—	—	—	—	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
8—9	—	—	—	—	—	1	6	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
9—10	—	—	—	—	—	—	3	3	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
10—11	—	—	—	—	—	—	2	6	3	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14
11—12	—	—	—	—	—	—	—	3	1	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
12—13	—	—	—	—	—	—	—	—	2	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
13—14	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
14—15	—	—	—	—	—	—	—	—	—	1	1	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	5
15—16	—	—	—	—	—	—	—	—	—	—	—	1	6	1	2	2	1	—	—	—	—	—	—	—	—	—	13
16—17	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	1	1	—	—	—	—	—	—	—	—	—	5
17—18	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	—	—	—	—	—	—	—	—	—	—	—	3
18—19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	1	—	—	—	—	—	—	—	—	—	3
19—20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	2
20—21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	2
21—22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	1	—	—	1	—	—	—	—	4
22—23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	—	—	—	—	—	3
23—24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
24—25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
25—26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
26—27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
27—28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
28—29	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
29—30	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
30—31	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	—	1	2	1	2	4	12	15	9	7	5	5	8	2	8	4	6	1	4	1	—	2	—	—	—	1	100

TABLE D.

χ'^2 from Basic Sample of 105 *and* χ^2 *from Parent Population.*

Parent Population χ^2

Basic Sample χ'^2	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12	12—13	13—14	14—15	15—16	16—17	17—18	18—19	19—20	20—21	21—22	22—23	23—24	24—25	25—26	26—27	27—28	Totals
2—3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3—4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
4—5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
5—6	—	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
6—7	—	—	—	—	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
7—8	—	—	—	—	1	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
8—9	—	—	—	—	—	2	6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
9—10	—	—	—	—	—	—	2	6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
10—11	—	—	—	—	—	1	1	7	5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14
11—12	—	—	—	—	—	—	—	2	1	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
12—13	—	—	—	—	—	—	—	—	1	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
13—14	—	—	—	—	—	—	—	—	2	—	1	—	3	—	—	—	—	—	—	—	—	—	—	—	—	—	6
14—15	—	—	—	—	—	—	—	—	—	1	1	1	2	—	1	—	—	—	—	—	—	—	—	—	—	—	6
15—16	—	—	—	—	—	—	—	—	—	—	1	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	4
16—17	—	—	—	—	—	—	—	—	—	—	—	3	—	1	1	—	1	—	—	—	—	—	—	—	—	—	6
17—18	—	—	—	—	—	—	—	—	—	—	—	—	2	1	1	—	1	—	—	—	—	—	—	—	—	—	5
18—19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	1	—	—	—	—	—	—	—	—	—	—	4
19—20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	2	—	—	—	—	—	—	—	—	—	—	3
20—21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	2
21—22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
22—23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	2	—	—	—	—	3
23—24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	2
24—25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	2
25—26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
26—27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	2
27—28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
28—29	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
29—30	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
30—31	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	—	1	2	1	2	4	12	15	9	7	5	5	8	2	8	4	6	1	4	1	—	2	—	—	—	1	100

TABLE E.

χ'^2 from Basic Sample of 60 and χ^2 from Parent Population.

Parent Population χ^2

Basic Sample χ'^2	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12	12—13	13—14	14—15	15—16	16—17	17—18	18—19	19—20	20—21	21—22	22—23	23—24	24—25	25—26	26—27	27—28	Totals
2—3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3—4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
4—5	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
5—6	—	—	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
6—7	—	—	—	—	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
7—8	—	—	—	—	—	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
8—9	—	—	—	—	—	1	5	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
9—10	—	—	—	—	—	—	5	3	4	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15
10—11	—	—	—	—	—	—	1	7	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11
11—12	—	—	—	—	—	—	—	2	1	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
12—13	—	—	—	—	—	—	—	—	2	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
13—14	—	—	—	—	—	—	—	—	—	1	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
14—15	—	—	—	—	—	—	—	—	—	1	—	1	1	—	1	2	—	—	—	—	—	—	—	—	—	—	6
15—16	—	—	—	—	—	—	—	—	—	—	—	1	6	1	1	—	1	—	—	—	—	—	—	—	—	—	10
16—17	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	1	—	—	—	—	—	—	—	—	—	—	3
17—18	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4	—	1	—	—	—	—	—	—	—	—	—	5
18—19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	1	—	—	—	—	—	—	—	—	3
19—20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	3	—	1	—	—	—	—	—	—	—	5
20—21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	2
21—22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	1	—	1	—	—	—	—	4
22—23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
23—24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
24—25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
25—26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
26—27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
27—28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
28—29	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
29—30	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
30—31	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	—	1	2	1	2	4	12	15	9	7	5	5	8	2	8	4	6	1	4	1	—	2	—	—	—	1	100

TABLE F.

χ'^2 from Basic Sample of 45 and χ^2 from Parent Population.

Parent Population χ^2

Basic Sample χ'^2	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12	12—13	13—14	14—15	15—16	16—17	17—18	18—19	19—20	20—21	21—22	22—23	23—24	24—25	25—26	26—27	27—28	Totals
2—3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3—4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
4—5	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
5—6	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
6—7	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
7—8	—	—	—	1	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
8—9	—	—	—	—	—	1	4	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
9—10	—	—	—	—	—	—	3	5	2	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11
10—11	—	—	—	—	—	—	5	3	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
11—12	—	—	—	—	—	—	—	5	3	5	1	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	17
12—13	—	—	—	—	—	—	—	—	1	1	1	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	5
13—14	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	2
14—15	—	—	—	—	—	—	—	—	—	—	2	2	1	—	3	—	—	—	—	—	—	—	—	—	—	—	8
15—16	—	—	—	—	—	—	—	—	—	—	—	—	2	—	1	—	1	—	—	—	—	—	—	—	—	—	4
16—17	—	—	—	—	—	—	—	—	—	—	—	—	2	1	1	3	1	—	2	—	—	—	—	—	—	—	10
17—18	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	1	1	—	1	—	—	—	—	—	—	5
18—19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	1	—	—	—	—	—	—	—	3
19—20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
20—21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	2	—	—	—	—	—	—	—	—	—	3
21—22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
22—23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
23—24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
24—25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1	2
25—26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
26—27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
27—28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Totals	—	1	2	1	2	4	12	15	9	7	5	5	8	2	8	4	6	1	4	1	—	2	—	—	—	1	100

TABLE G.

χ'^2 from Basic Sample of 30 and χ^2 from Parent Population.

Parent Population χ^2

Basic Sample χ'^2	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12	12—13	13—14	14—15	15—16	16—17	17—18	18—19	19—20	20—21	21—22	22—23	23—24	24—25	25—26	26—27	27—28	Totals
2—3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3—4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
4—5	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
5—6	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
6—7	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
7—8	—	—	—	—	1	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
8—9	—	—	—	—	1	1	9	5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	16
9—10	—	—	—	—	—	2	2	6	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11
10—11	—	—	—	—	—	—	—	1	1	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	4
11—12	—	—	—	—	—	—	—	2	3	2	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
12—13	—	—	—	—	—	—	—	—	1	1	1	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	5
13—14	—	—	—	—	—	—	—	—	2	3	—	1	3	—	—	—	—	—	—	—	—	—	—	—	—	—	9
14—15	—	—	—	—	—	—	—	—	—	—	1	2	1	—	1	—	—	—	—	—	—	—	—	—	—	—	5
15—16	—	—	—	—	—	—	—	—	—	—	—	—	1	—	3	—	—	—	—	—	—	—	—	—	—	—	4
16—17	—	—	—	—	—	—	—	—	—	—	—	—	—	1	2	2	—	—	—	—	—	—	—	—	—	—	5
17—18	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	3	—	—	—	—	—	—	—	—	—	5
18—19	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	2	—	—	—	—	—	—	—	—	—	3
19—20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1	—	—	—	—	—	—	—	2
20—21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	1	—	—	—	—	—	—	—	3
21—22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
22—23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	1	—	—	—	—	3
23—24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	2
24—25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
25—26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
26—27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1	2
27—28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Totals	—	1	2	1	2	4	12	15	9	7	5	5	8	2	8	4	6	1	4	1	—	2	—	—	—	1	100

TABLE H.

χ'^2 from Basic Sample of 15 and χ^2 from Parent Population.

Parent Population χ^2

Basic Sample χ'^2	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12	12—13	13—14	14—15	15—16	16—17	17—18	18—19	19—20	20—21	21—22	22—23	23—24	24—25	25—26	26—27	27—28	Totals
2—3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3—4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
4—5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
5—6	—	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
6—7	—	—	—	—	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
7—8	—	—	—	—	—	—	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
8—9	—	—	—	—	—	2	4	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
9—10	—	—	—	—	—	—	3	6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
10—11	—	—	—	—	—	1	1	2	3	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
11—12	—	—	—	—	—	—	—	4	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
12—13	—	—	—	—	—	—	—	1	1	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
13—14	—	—	—	—	—	—	—	—	3	1	1	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	7
14—15	—	—	—	—	—	—	—	—	—	—	2	1	2	—	1	—	—	—	—	—	—	—	—	—	—	—	6
15—16	—	—	—	—	—	—	—	—	—	1	—	2	2	1	—	—	—	—	—	—	—	—	—	—	—	—	6
16—17	—	—	—	—	—	—	—	—	—	—	—	1	—	—	2	—	1	—	—	—	—	—	—	—	—	—	4
17—18	—	—	—	—	—	—	—	—	—	—	—	1	1	1	—	—	1	—	—	—	—	—	—	—	—	—	4
18—19	—	—	—	—	—	—	—	—	—	—	—	—	1	—	2	2	—	—	—	—	—	—	—	—	—	—	5
19—20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	—	—	—	—	—	—	—	—	—	3
20—21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	2
21—22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	2
22—23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	2	—	—	—	—	3
23—24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
24—25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
25—26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	2
26—27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
27—28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	2
28—29	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
29—30	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
30—31	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
31—32	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
32—33	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	—	1	2	1	2	4	12	15	9	7	5	5	8	2	8	4	6	1	4	1	—	2	—	—	—	1	100

ON THE DISTRIBUTION OF THE CORRELATION COEFFICIENT IN SMALL SAMPLES*.

By PAUL R. RIDER, Washington University, Saint Louis.

It was the original purpose of this study to attempt to discover the effect upon the distribution in random samples, particularly in small samples, of the product-moment coefficient of correlation, r, when the samples are drawn from a non-normal instead of a normal population. In Part I the results of sampling from certain populations which differ greatly from the normal are given, also the results of sampling from a normal population having a high degree of correlation. As the sampling was done experimentally, it was necessary to deal with discrete populations. This opened up the question of the effect of grouping upon the distribution of r, a question which is investigated in Part II.

I. Sampling from Non-Normal Populations and from a Normal Population having High Correlation.

Description of the Populations sampled.

The populations sampled will be termed rectangular, triangular, and normal, and will be designated by R, T, and N, respectively.

Rectangular Population. The rectangular population may best be characterized by its correlation table, which is composed of 10×10 compartments, each with the same frequency. That is, there is equal probability, in random sampling, of obtaining any pair of values within a limited (rectangular) region. Obviously the correlation, ρ, in such a population is zero. If x and y are the variates, then for the marginal distributions we have $\beta_1(x) = \beta_1(y) = 0$, $\beta_2(x) = \beta_2(y) = 1{\cdot}77\dot{5}\dot{7}$. (The dots indicate a repeating decimal.) For the corresponding continuous bivariate distribution, $\rho = 0$, $\beta_1(x) = \beta_1(y) = 0$, $\beta_2(x) = \beta_2(y) = 1{\cdot}8$. The frequency surface is a rectangular parallelopiped, or, with suitably chosen units, a cube.

Triangular Population. The correlation table of the triangular population is also composed of 10×10 cells, but the frequencies in all of the cells on one side of a principal diagonal are zeros, the frequencies in the remaining cells having a constant value different from zero. It is found that

$$\rho = \tfrac{1}{2}, \quad \beta_1(x) = \beta_1(y) = 0{\cdot}32\dot{6}, \quad \beta_2(x) = \beta_2(y) = 2{\cdot}3\dot{6}.$$

* This investigation was made possible by assistance from a grant made by the *Rockefeller Foundation* to Washington University for Research in Science. The writer wishes to express his sincere thanks for this grant, and also to make grateful acknowledgment of valuable criticisms, suggestions, and assistance given by Dr Egon S. Pearson.

For the corresponding continuous distribution the frequency surface is a right prism, with a right triangle for a base. Its constants are

$$\rho = \tfrac{1}{2},\quad \beta_1(x) = \beta_1(y) = 0{\cdot}32,\quad \beta_2(x) = \beta_2(y) = 2{\cdot}4.$$

Normal Population. The normal population sampled is the one shown in Table I. Its frequencies were calculated from tables of volumes of the normal surface*. It

TABLE I.

Normal Correlation Table.

								Totals
	—	—	—	—	1	29	32	62
	—	—	—	5	223	349	29	606
	—	—	8	618	1566	223	1	2416
	—	5	618	2582	618	5	—	3828
	1	223	1566	618	8	—	—	2416
	29	349	223	5	—	—	—	606
	32	29	1	—	—	—	—	62
Totals	62	606	2416	3828	2416	606	62	9996

$\rho = 0{\cdot}9$ with Sheppard's correction.
$\rho = 0{\cdot}83$ without Sheppard's correction.

represents a population in which the coefficient of correlation is 0·9. On account of the grouping, the actual coefficient as computed from the table is 0·83, but when Sheppard's correction is applied to the two standard deviations involved in the denominator of the coefficient of correlation, its value is 0·9009. The values of β_1 and β_2 for the marginal distributions are 0 and 2·9390 respectively. With Sheppard's corrections, $\beta_2 = 2{\cdot}9367$.

It was suggested by Karl Pearson, in a letter to the writer, that it would be desirable to test by actual sampling, whether observed values of r actually follow the theoretical distribution when there is high correlation and *grouped* frequencies in the sampled population. It was the object of this particular experiment to test the theory.

Results of Sampling.

The sampling was effected by the use of Tippett's numbers†. One thousand samples of 5 pairs each were obtained from each of the populations R, T, and N. By clubbing these together in the case of T and of N, five hundred samples of 10 pairs each were obtained from each of these two populations. The observed distributions of r are given in Table II.

* See *Tables for Statisticians and Biometricians*, Part II, Table VIII, pp. 105—106.

† L. H. C. Tippett, *Random Sampling Numbers* (Cambridge University Press, Tracts for Computers, No. 15).

TABLE II.

Distribution of r in Samples of 5 *and of* 10.

r	Frequencies				
	R_5	T_5	T_{10}	N_5	N_{10}
·95 to 1·00	9	48	3	235	34
·90 „ ·95	10	59	9	175·5	97
·85 „ ·90	17	49	11	146·5	92
·80 „ ·85	20	62	17	83	86
·75 „ ·80	19	64	28	92·5	46
·70 „ ·75	22	59	27	33·5	56
·65 „ ·70	35	64	40	36	23
·60 „ ·65	21	47	46	65	21
·55 „ ·60	33	55	54	12	15
·50 „ ·55	23	46	31	27	15
·45 „ ·50	23	53	36	14	4
·40 „ ·45	27	50	36	21	3
·35 „ ·40	33	38	38	2	3
·30 „ ·35	30	25	19	14	1
·25 „ ·30	27	40	22	4	—
·20 „ ·25	26	27	22	4	2
·15 „ ·20	33	31	15	8	—
·10 „ ·15	32	14	22	4	—
·05 „ ·10	33	20	6	1	—
·00 „ ·05	24	16·5	2·5	4	—
·00 „ −·05	33	17·5	4·5	4	—
−·05 „ −·10	29	15	2	—	1
−·10 „ −·15	32	4	1	—	1
−·15 „ −·20	28	9	—	4	—
−·20 „ −·25	29	9·5	3	4	—
−·25 „ −·30	28	11·5	3	2	—
−·30 „ −·35	24	11	—	—	—
−·35 „ −·40	28	5	—	—	—
−·40 „ −·45	27	8	—	4	—
−·45 „ −·50	41	6	—	—	—
−·50 „ −·55	24	5	—	—	—
−·55 „ −·60	28	6	—	—	—
−·60 „ −·65	29	4	1	—	—
−·65 „ −·70	21	2	—	—	—
−·70 „ −·75	21	5·5	—	—	—
−·75 „ −·80	16	6·5	—	—	—
−·80 „ −·85	17	3	—	—	—
−·85 „ −·90	27	1	—	—	—
−·90 „ −·95	9	1	1	—	—
−·95 „ −1·00	12	2	—	—	—
Total	1000	1000	500	1000	500

R_5 refers to samples of 5 from rectangular population R.
T_5 and T_{10} refer to samples of 5 and of 10 respectively from triangular population T.
N_5 and N_{10} refer to samples of 5 and of 10 respectively from normal population N.

Pearson curves were fitted to three of these distributions with the following results: for 1000 samples of 5 from the rectangular population, a Type II curve*

$$y = 1000 \times 0{\cdot}60632\,(1 - r^2)^{0{\cdot}37954} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1);$$

for 1000 samples of 5 from the triangular population, a Type I curve

$$y = 1000 \times 0{\cdot}11157\,(1{\cdot}54516 + r)^{2{\cdot}6895}\,(0{\cdot}92826 - r)^{-0{\cdot}16416} \quad \ldots\ldots\ldots(2);$$

for 500 samples of 10 from the triangular population, a Type IV curve

$$y = 500 \times 0{\cdot}19841\left(1 + \frac{x^2}{0{\cdot}57672}\right)^{-5{\cdot}80845} e^{-7{\cdot}4102\tan^{-1}(x/0{\cdot}5767)} \quad \ldots\ldots\ldots(3);$$

in which $x = r - 0{\cdot}9653$.

In cases (2) and (3) the fit was effected by equating the first four moments of the curve to the corresponding moments of the observations (without correcting for abruptness). In fitting the Type II curve to the distribution of r in 1000 samples of 5 from the rectangular population it was found that

$$\bar{r} = 0{\cdot}04685,\quad \mu_2 = 0{\cdot}2660225,\quad \mu_3 = -0{\cdot}0001043661625,$$

$$\mu_4 = 0{\cdot}13714900615625,\quad \beta_1 = 0{\cdot}0005817,\quad \beta_2 = 1{\cdot}944985,$$

$$a = \left(\frac{2\mu_2\beta_2}{3 - \beta_2}\right)^{\frac{1}{2}} = 0{\cdot}9895,\quad m = \frac{5\beta_2 - 9}{2(3 - \beta_2)} = 0{\cdot}34356\dagger.$$

Assuming that the small value of β_1 justifies a Type II curve, we should get

$$y = 1000 \times 0{\cdot}596989\left[1 - \left(\frac{r - 0{\cdot}04685}{0{\cdot}9895}\right)^2\right]^{0{\cdot}34356} \quad \ldots\ldots\ldots\ldots(1a).$$

It seemed more desirable, however, to fit a curve of the type $y = y_0(1 - r^2)^k$, determining k and y_0 so that theoretical and observed values are equal for standard deviations and totals. It is readily found that

$$k = \frac{1}{2}\left(\frac{1}{\mu_2} - 3\right),\quad y_0 = \frac{N\Gamma(2k+2)}{2^{2k+1}[\Gamma(k+1)]^2} \text{ or } \frac{N\Gamma(k + 3/2)}{\sqrt{\pi}\,\Gamma(k+1)},$$

N being the total number of values of r, that is, the number of samples. Equation (1) was derived in this fashion.

Although the exact distribution of r for samples of size n from a normal population has been given ‡, the general distribution equation is somewhat complicated, except in the case when $\rho = 0$. In the latter case for the distribution of 1000 samples of 5 it becomes

$$y = 1000 \,.\, \frac{2}{\pi}(1 - r^2)^{\frac{1}{2}} = 636{\cdot}62\,(1 - r^2)^{\frac{1}{2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(4),$$

an equation which should be compared with (1).

A graphical comparison of results is made in Figures 1—6. The graphs of the theoretical distributions of r for samples from continuous normal populations were plotted from tables of ordinates of the frequency curves of the correlation coefficient§.

* See next paragraph.

† For notation see Elderton: *Frequency Curves and Correlation.*

‡ R. A. Fisher: "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population." *Biometrika*, Vol. x (1914—15), pp. 507—521.

§ *Biometrika*, Vol. xi (1915—17), pp. 379 ff.

In the case of samples of 10 from a normal population it will be recalled that the value of ρ, if Sheppard's correction be not applied, is 0·83. The value of ρ in the corresponding continuous population is 0·9. The theoretical distribution curves for both $\rho = 0{\cdot}8$ and $\rho = 0{\cdot}9$ are plotted. The histogram representing the actual samples

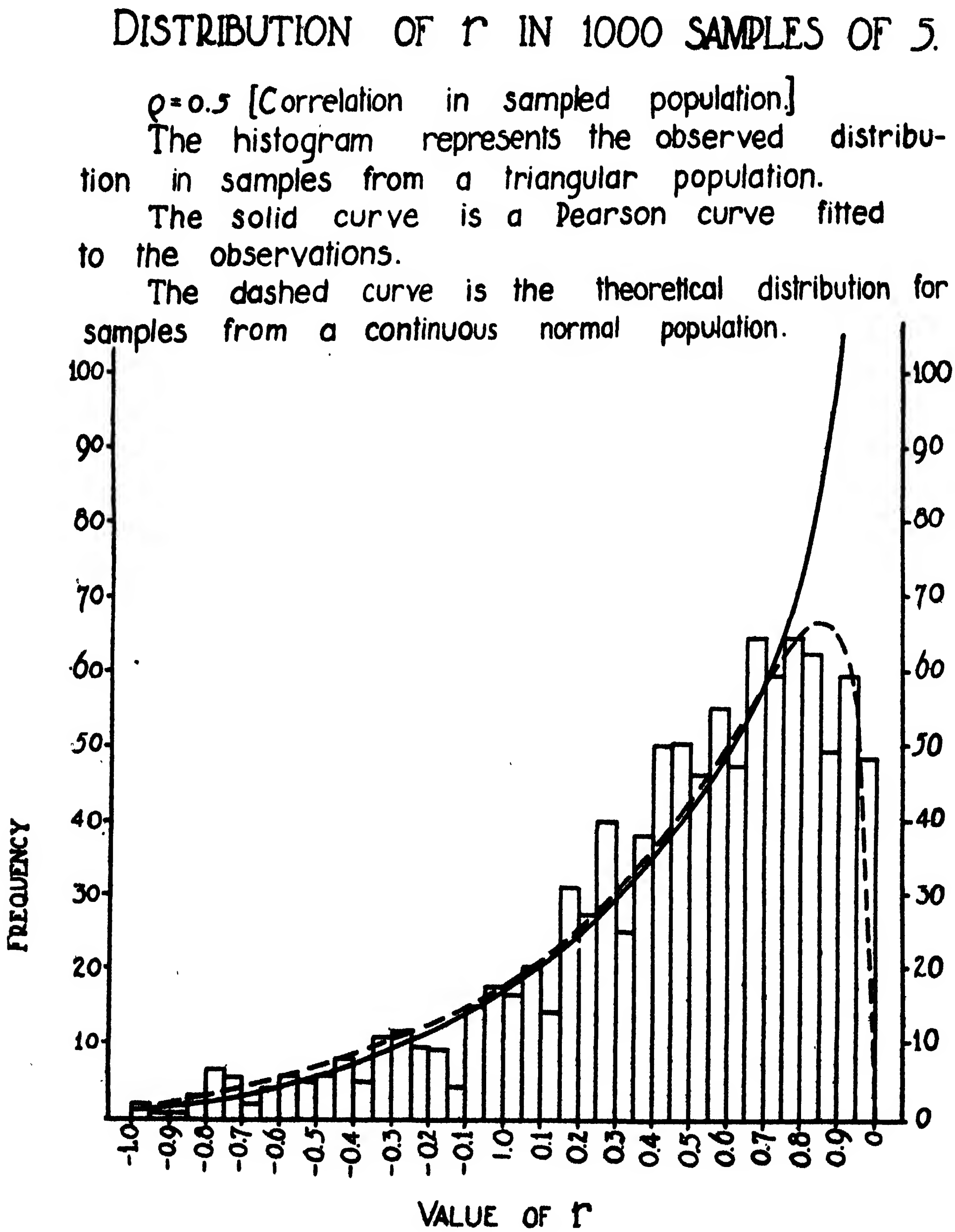


Fig. 1.

seems to lie closer to the curve corresponding to $\rho = 0{\cdot}8$, and it appears that an even better fit would be obtained if a curve for $\rho = 0{\cdot}83$ were interpolated between the two given curves.

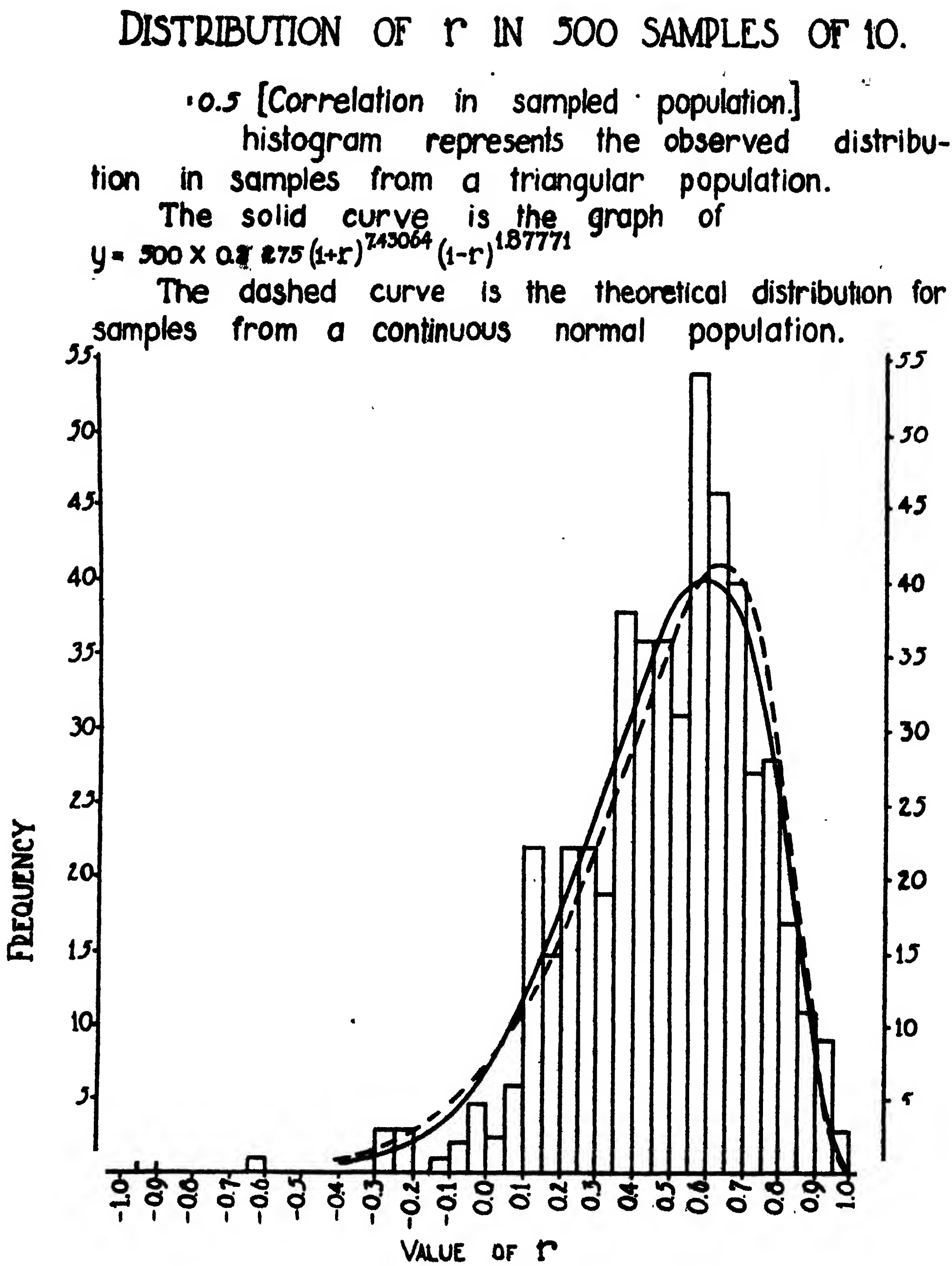


Fig. 2.

It will be noted that the Pearson curves fitted to the values of r from the samples drawn from the triangular population fail to have the proper range. It was consequently thought desirable to fit a curve of the type $y = y_0(1+r)^{k_1}(1-r)^{k_2}$. It

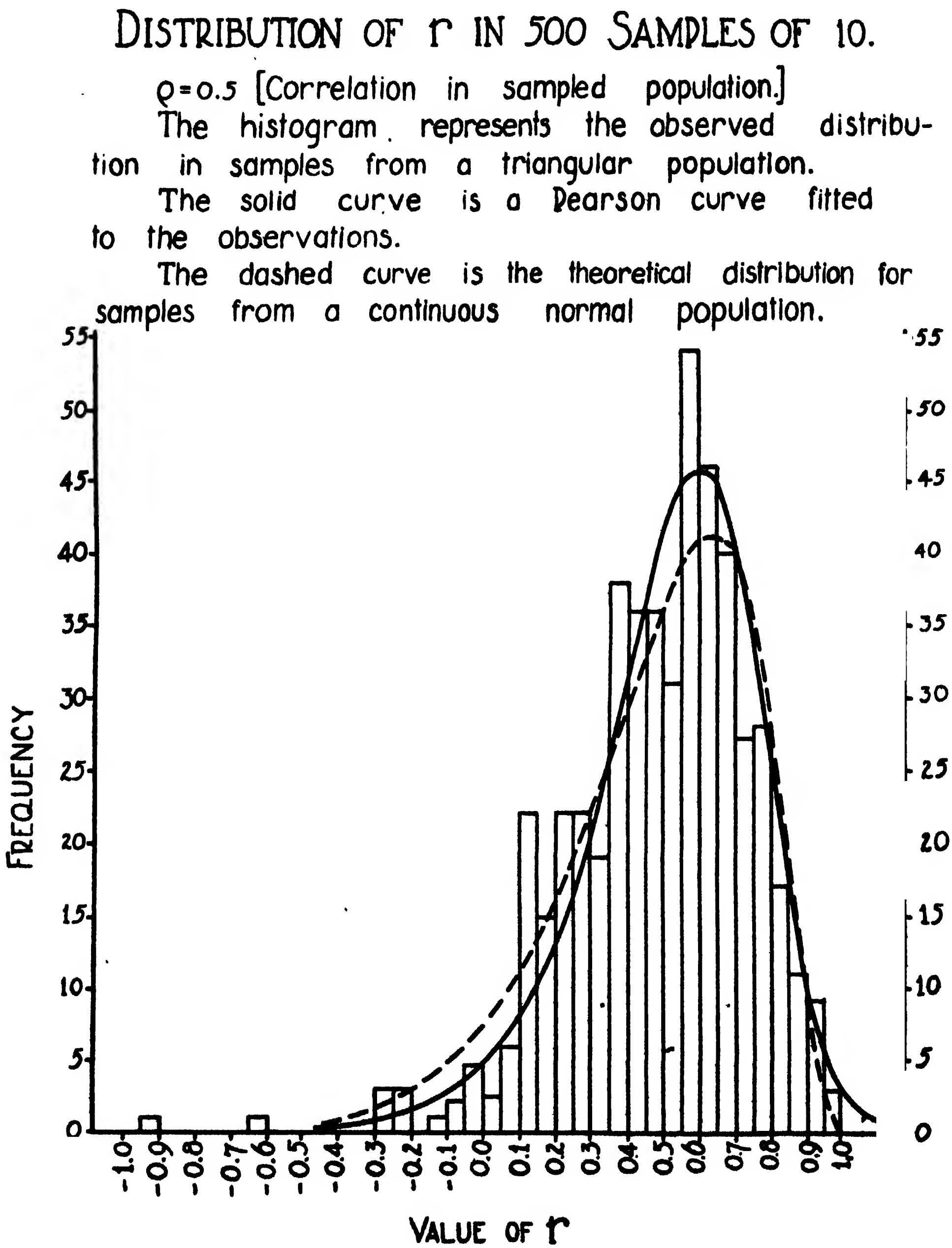


Fig. 8.

is found that if the fitting is done by the method of moments the constants have the following values:

$$k_1 = \frac{(1+\bar{r})^2(1-\bar{r})-(\bar{r}+3)\sigma_r^2}{2\sigma_r^2},$$

$$k_2 = \frac{(1-\bar{r})^2(1+\bar{r})+(\bar{r}-3)\sigma_r^2}{2\sigma_r^2},$$

$$y_0 = \frac{N\Gamma(k_1+k_2+2)}{2^{k_1+k_2+1}\Gamma(k_1+1)\Gamma(k_2+1)},$$

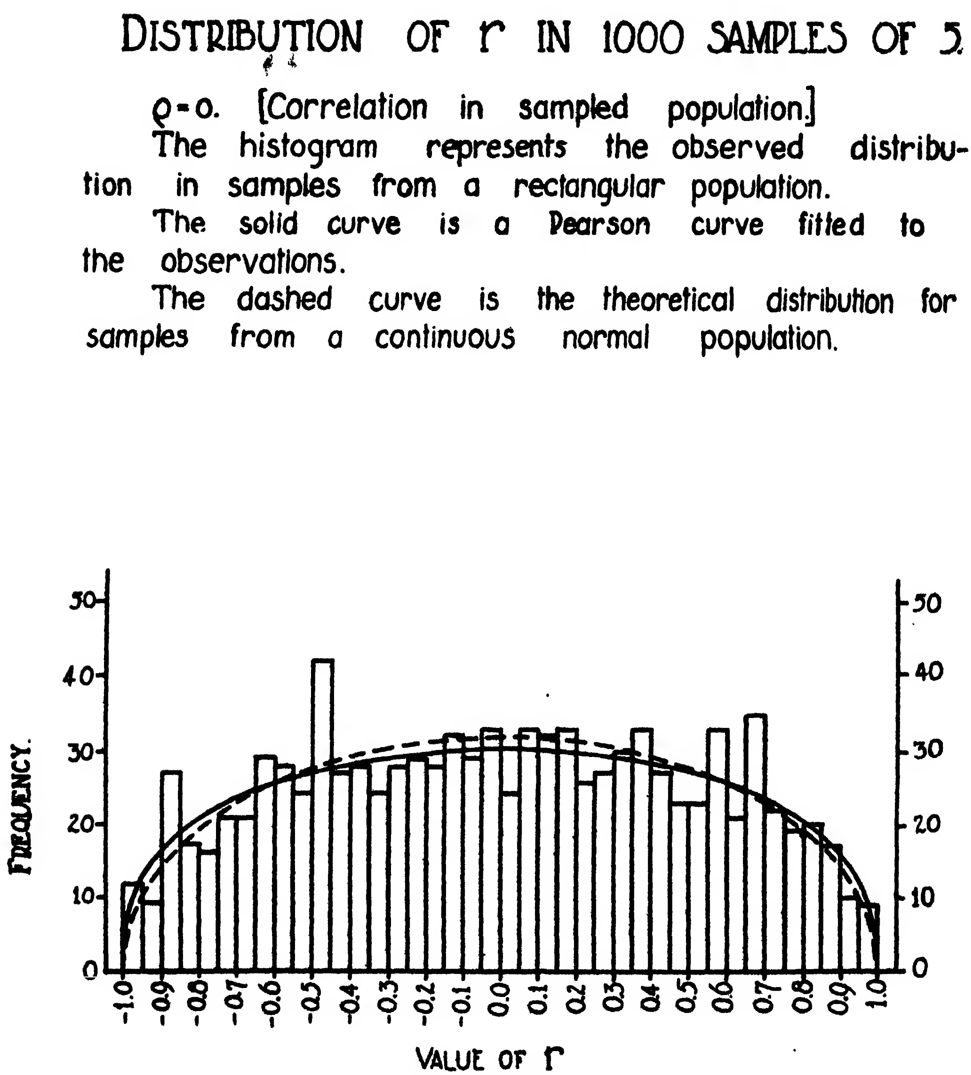


Fig. 4.

in which N is the number of samples. For the samples of 10 the equation of the curve is

$$y = 500 \times 0{\cdot}27275\,(1+r)^{7{\cdot}48064}(1-r)^{1{\cdot}87771} \quad \text{.................(5).}$$

The graph is shown in Figure 1.

TABLE III.

Goodness of Fit. Samples of 5 from Rectangular Population. ($\rho = 0.$)

r	Observed frequency f_0	Fitted Pearson curve, f_1	"Normal theory" frequency, f_2	$\frac{(f_0-f_1)^2}{f_1}$	$\frac{(f_0-f_2)^2}{f_2}$
·95 to 1·00	9	9·0	6·2	·0000	·7896
·90 ,, ·95	10	14·8	11·9	1·5568	·3034
·85 ,, ·90	17	17·4	15·4	·0092	·1662
·80 ,, ·85	20	19·6	18·0	·0082	·2222
·75 ,, ·80	19	21·4	20·1	·2692	·0602
·70 ,, ·75	22	22·8	21·9	·0105	·0005
·65 ,, ·70	35	24·0	23·4	5·0417	5·7504
·60 ,, ·65	21	25·1	24·9	·6697	·6108
·55 ,, ·60	33	26·0	26·0	1·8846	1·8846
·50 ,, ·55	23	26·8	27·1	·5388	·6203
·45 ,, ·50	23	27·5–	28·0	·7364	·8929
·40 ,, ·45	27	28·1	28·7	·0431	·1007
·35 ,, ·40	33	28·6	29·6	·6769	·3905
·30 ,, ·35	30	29·0	30·1	·0345	·0003
·25 ,, ·30	27	29·4	30·5–	·1959	·4016
·20 ,, ·25	26	29·7	31·1	·4609	·8363
·15 ,, ·20	33	30·0	31·3	·3000	·0923
·10 ,, ·15	32	30·1	31·6	·1199	·0051
·05 ,, ·10	33	30·2	31·7	·2596	·0531
·00 ,, ·05	24	30·3	31·8	1·3099	1·9132
·00 ,, –·05	33	30·3	31·8	·2406	·0453
–·05 ,, –·10	29	30·2	31·7	·0477	·2230
–·10 ,, –·15	32	30·1	31·6	·1199	·0051
–·15 ,, –·20	28	30·0	31·3	·1333	·3479
–·20 ,, –·25	29	29·7	31·1	·0165	·1418
–·25 ,, –·30	28	29·4	30·5–	·0667	·2049
–·30 ,, –·35	24	29·0	30·1	·8621	1·2362
–·35 ,, –·40	28	28·6	29·6	·0126	·0865
–·40 ,, –·45	27	28·1	28·7	·0431	·1007
–·45 ,, –·50	41	27·5–	28·0	6·6273	6·0357
–·50 ,, –·55	24	26·8	27·1	·2925	·3546
–·55 ,, –·60	28	26·0	26·0	·1538	·1538
–·60 ,, –·65	29	25·1	24·9	·6060	·6751
–·65 ,, –·70	21	24·0	23·4	·3750	·2462
–·70 ,, –·75	21	22·8	21·9	·1421	·0370
–·75 ,, –·80	16	21·4	20·1	1·3626	·8363
–·80 ,, –·85	17	19·6	18·0	·3449	·0556
–·85 ,, –·90	27	17·4	15·4	5·2966	8·7377
–·90 ,, –·95	9	14·8	11·9	2·2730	·7067
–·95 ,, –1·00	12	9·0	6·7	1·0000	4·1925
Totals	1000	999·6	999·6	34·1421	39·5168

For fitted Pearson curve, $\chi^2 = 34{\cdot}1421$, $n = 38$, $P = 0{\cdot}648$.
For "normal theory" frequency, $\chi^2 = 39{\cdot}5168$, $n = 39$, $P = 0{\cdot}448$.

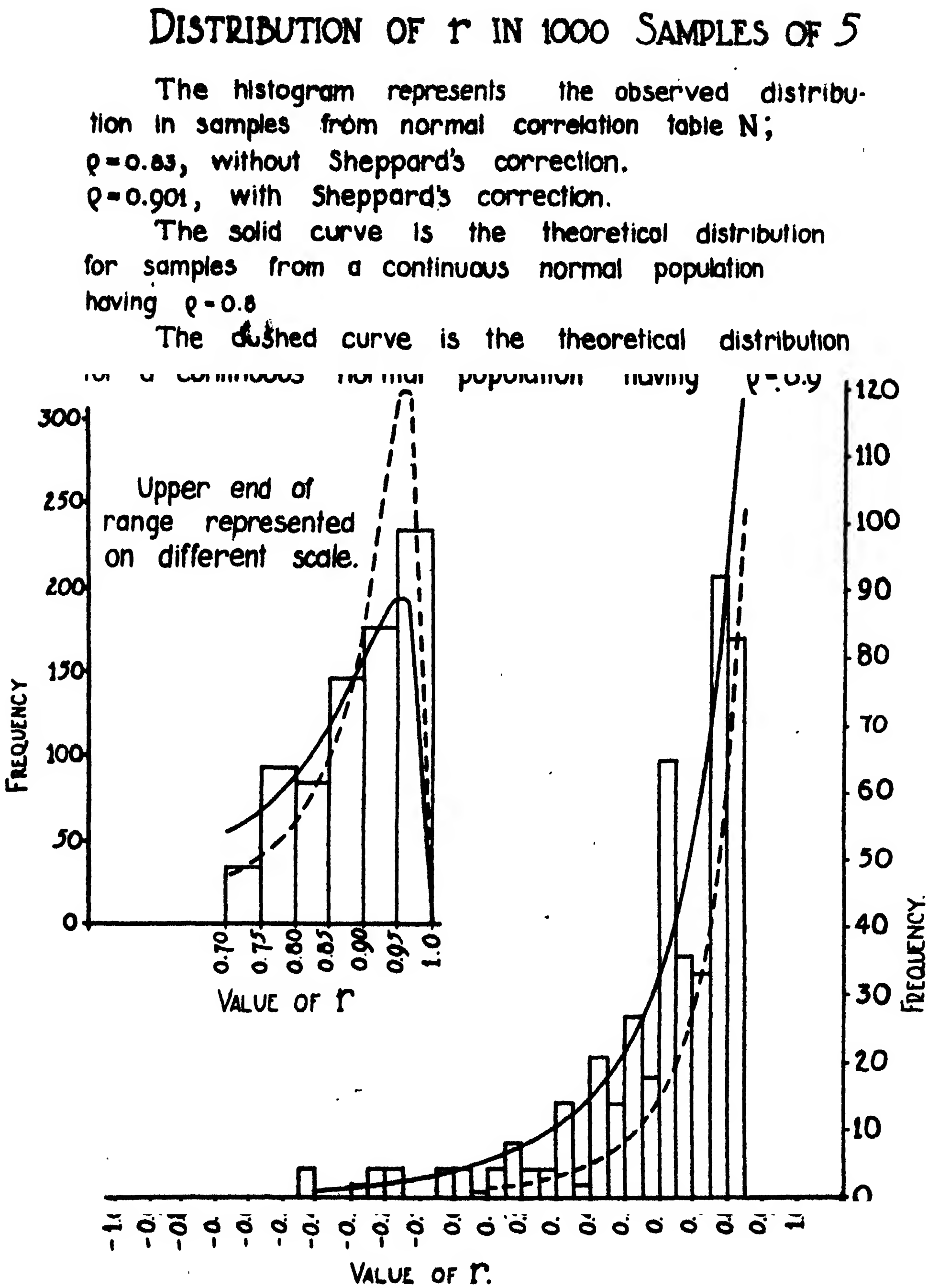

DISTRIBUTION OF r IN 1000 SAMPLES OF 5
The histogram represents the observed distribution in samples from normal correlation table N;
ρ=0.83, without Sheppard's correction.
ρ=0.901, with Sheppard's correction.
The solid curve is the theoretical distribution for samples from a continuous normal population having ρ=0.8
The dashed curve is the theoretical distribution
Upper end of range represented on different scale.
FREQUENCY
300
250
200
150
100
50
0
0.70
0.75
0.80
0.85
0.90
0.95
1.0
VALUE OF r
120
110
100
90
80
70
60
50
40
30
20
10
0
FREQUENCY
VALUE OF r.

Fig. 5.

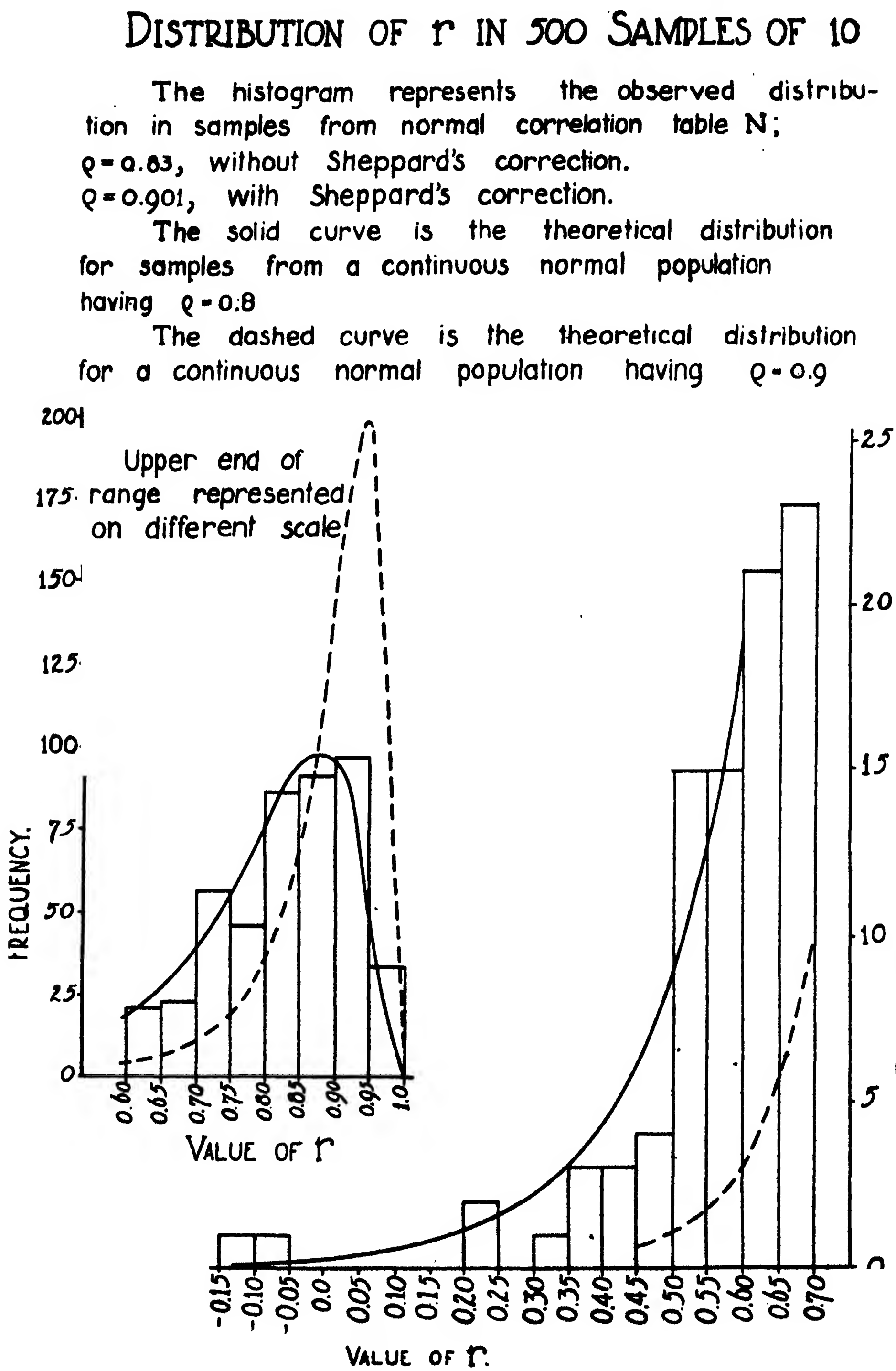

DISTRIBUTION OF r IN 500 SAMPLES OF 10
The histogram represents the observed distribution in samples from normal correlation table N;
ρ = 0.83, without Sheppard's correction.
ρ = 0.901, with Sheppard's correction.
The solid curve is the theoretical distribution for samples from a continuous normal population having ρ = 0.8
The dashed curve is the theoretical distribution for a continuous normal population having ρ = 0.9
Upper end of range represented on different scale
FREQUENCY
200
175
150
125
100
75
50
25
0
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.0
VALUE OF r
25
20
15
10
5
0
FREQUENCY
-0.15
-0.10
-0.05
0.0
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70
VALUE OF r.

Fig. 6.

The corresponding equation for samples of 5 was not worked out, but it was noted that the value of k_2 was negative, yielding a curve very much like the solid curve of Figure 2, starting, however, at $r=-1$ and having an asymptote at $r=1$.

Goodness of Fit.

The computation, especially the mechanical quadrature, necessary to apply the χ^2 goodness of fit test to all of the distributions did not seem warranted. However, the test was applied in several instances.

The application to the distribution of r in the samples of 5 from the rectangular population (the case represented in Figure 3) is shown in Table III. The fitted Pearson curve is given by (1); the "normal theory" frequencies corresponding to samples of 5 from a normal population in which the correlation ρ is zero were obtained from Fisher's curve (4). For the Pearson curve, $\chi^2=34\cdot1421$, and since there are 40 groups and the theoretical distribution has been made to agree with the observed in the total and in the standard deviation $n=38$, using the notation $n+1=n'$ of Elderton's Table (Table XII of Pearson's *Tables for Statisticians and Biometricians*, Part I). These values are beyond the range of this Table, but by means of *Tables of the Incomplete* Γ*-function* it is found that $P=0\cdot648$. (R. A. Fisher's

TABLE IV.

Goodness of Fit. Samples of 5 from Triangular Population. ($\rho=0\cdot5$.)

r	Observed frequency f_0	"Normal theory" frequency, f	$\frac{(f_0-f)^2}{f}$
·9 to 1·0	107	98·9	·66
·8 „ ·9	111	131·8	3·28
·7 „ ·8	123	124·2	·01
·6 „ ·7	111	109·0	·04
·5 „ ·6	101	93·2	·65
·4 „ ·5	103	78·8	7·43
·3 „ ·4	63	66·3	·16
·2 „ ·3	67	55·7	2·29
·1 „ ·2	45	46·7	·06
0 „ ·1	36·5	39·2	·19
0 „ −·1	32·5	32·8	·00
−·1 „ −·2	13	27·4	5·54
−·2 „ −·3	21	22·9	·16
−·3 „ −·4	16	19·0	·47
−·4 „ −·5	14	15·6	·16
−·5 „ −·6	11	12·7	·23
−·6 „ −·7	6	10·1	1·66
−·7 „ −·8	12	7·7	2·40
−·8 „ −·9	4	5·4	·12
−·9 „ −1·0	3	2·6	
Totals	1000·0	1000·0	25·51

$\chi^2=25\cdot51$, $n=18$, $P=0\cdot112$.

approximate method* yields the value $P = 0{\cdot}654$.) For the Fisher curve for samples of 5 from uncorrelated normal material, $\chi^2 = 39{\cdot}5168$, $n = 39$ (since there are 40 classes and the theoretical distribution is made to agree only in the total), $P = 0{\cdot}448$. (Fisher's approximate method gives the value $P = 0{\cdot}454$.)

The goodness of fit test for samples of 5 from the triangular population is shown in Table IV. The "normal theory" frequencies are those corresponding to samples of 5 from a normal population in which $\rho = 0{\cdot}5$ †.

Here $\chi^2 = 25{\cdot}51$, $n = 18$ ($n' = 19$ for use in Pearson's *Tables for Statisticians and Biometricians*), and $P = 0{\cdot}112$.

For samples of 10 from the triangular population (see Table V) the value of χ^2 is 46·89, $n = 22$, and $P = 0{\cdot}00153$, the only extremely bad fit noted.

TABLE V.

Goodness of Fit. Samples of 10 *from Triangular Population.* ($\rho = 0{\cdot}5$.)

r	Observed frequency f_0	"Normal theory" frequency, f	$\frac{(f_0-f)^2}{f}$
·95 to 1·00	3	0·7	8·47
·90 „ ·95	9	4·6	
·85 „ ·90	11	12·9	·28
·80 „ ·85	17	22·4	1·30
·75 „ ·80	28	30·5	·20
·70 „ ·75	27	36·8	2·61
·65 „ ·70	40	40·1	·00
·60 „ ·65	46	41·1	·58
·55 „ ·60	54	40·2	4·74
·50 „ ·55	31	38·0	1·29
·45 „ ·50	36	34·9	·03
·40 „ ·45	36	31·5	·64
·35 „ ·40	38	27·9	3·66
·30 „ ·35	19	24·3	1·16
·25 „ ·30	22	20·9	·06
·20 „ ·25	22	17·7	1·04
·15 „ ·20	15	14·9	·00
·10 „ ·15	22	12·4	7·43
·05 „ ·10	6	10·2	1·73
·00 „ ·05	2·5	8·4	4·14
·00 „ −·05	4·5	6·8	·78
−·05 „ −·10	2	5·4	4·63
−·10 „ −·15	1	4·3	
−·15 „ −·20		3·4	1·50
−·20 „ −·25	3	2·6	
−·25 „ −·30	3	2·0	·62
−·30 „ −·35		1·5	
Below −·35	2	3·6	
Totals	500·0	500·0	46·89

$\chi^2 = 46{\cdot}89$, $n = 22$, $P = 0{\cdot}00153$.

* Put $t = \sqrt{2\chi^2} - \sqrt{2n-1}$; then $P = \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$ approximately.

† See *Biometrika*, Vol. XI, p. 381. The frequencies were obtained from the ordinates by quadrature.

The distribution of r for samples of 10 from the normal correlation table (Table I) in which $\rho = 0{\cdot}83$ before Sheppard's correction is applied ($\rho = 0{\cdot}9$ after) was tested for goodness of fit after making Fisher's transformation* $z = \tanh^{-1} r$, $\zeta = \tanh^{-1} \rho$. The variate z is then approximately normally distributed with mean

$$\bar{z} = \zeta + \frac{\rho}{2(n-1)}\left[1 + \frac{1+\rho^2}{8(n-1)} + \ldots\right] \qquad \text{.....................(6)}$$

and variance

$$\sigma_z^2 = \frac{1}{n-1}\left[1 + \frac{4-\rho^2}{2(n-1)} + \frac{176 - 21\rho^2 - 21\rho^4}{48(n-1)^2} + \ldots\right] \qquad \text{............(7).}$$

In the present instance, it is found that $\bar{z} = 1{\cdot}23533$, $\sigma_z^2 = 0{\cdot}13588$. The transformation and the χ^2 test are worked out in Table VI. It is found that $P = 0{\cdot}024$. The fit is not very good, but the main discrepancies are somewhat irregularly scattered throughout the distribution.

TABLE VI.

Goodness of Fit. Samples of 10 from Normal Population having High Correlation. Fisher's Transformation.

r	$z = \tanh^{-1} r$	$\frac{z-\bar{z}}{\sigma_z}$	Normal area $\frac{1}{2}(1+\alpha)$	Δ	$500\Delta = f$	Observed frequency f_0	$\frac{(f_0 - f)^2}{f}$
1·00	∞	∞	1·00000	·05281	26·41	34	2·19
·95	1·831 7808	+1·6182	·94719	·20741	103·71	97	·43
·90	1·472 2195	+ ·6427	·73978	·21725	108·62	92	2·54
·85	1·256 1528	+ ·0565	·52253	·16717†	83·59	86	·07
·80	1·098 6123	− ·3709	·64464	·11707	58·53	46	2·67
·75	·972 9551	− ·7118	·76171	7927	39·64	56	6·79
·70	·867 3005	− ·9985	·84098	5302	26·51	23	·46
·65	·775 2987	−1·2481	·89400	3534	17·67	21	·77
·60	·693 1472	−1·4709	·92934	2357	11·78	15	·87
·55	·618 3813	−1·6738	·95291	1573	7·87	15	6·38
·50	·549 3061	−1·8612	·96864	1050	5·25	4	·27
·45	·484 7003	−2·0365	·97914	703	3·51	3	·00
·40	·423 6489	−2·2021	·98617	469	2·35	3	
·35	·365 4438	−2·3600	·99086	313	1·56	1	
·30	·309 5196	−2·5117	·99399	209	1·05		
·25	·255 4128	−2·6585	·99608	137	·68	2	
·20	·202 7326	−2·8014	·99745	92	·46		·03
·15	·151 1404	−2·9414	·99837	60	·30		
·10	·100 3353	−3·0792	·99897	38	·19		
·05	·050 0417	−3·2157	·99935	25	·12		
·00	·000 0000	−3·3515	·99960	·00040	·20	(<0) 2	
−·05	−·050 0417						
					500·00	500	23·47

$\bar{z} = 1{\cdot}2351$, $\sigma_z = 0{\cdot}3686$, $\chi^2 = 23{\cdot}47$, $n = 12$, $P = 0{\cdot}024$.

* See R. A. Fisher, "On the 'probable error' of a Coefficient of Correlation deduced from a small Sample," *Metron*, Vol. I. No. 4 (Sept. 1, 1921), pp. 3—32. E. S. Pearson has given an illustration showing the degree of accuracy of this approximation in a very similar case ($n = 10$, $\rho = {\cdot}8$) (*Biometrika*, Vol. XXI, p. 359).

† Note that we have a positive and a negative value of $(z - \bar{z})/\sigma_z$.

Comparison of Constants.

In Table VII the means, the standard deviations, and the betas of the distributions of r for the observed samples from the various populations are compared with the respective constants* for the corresponding theoretical distributions for samples from normal populations.

TABLE VII.

Comparison of Constants for Distribution of Correlation Coefficient r.

Sampling Experiment		$\bar{r}$	σ_r	β_1	β_2
1000 samples of 5 from rectangular population ($\rho=0\cdot0$)	Experiment	$-0\cdot0031$	0·5153	0·0000 5817 ($\sqrt{\beta_1}=0\cdot0076$)	1·9450⁻
	Normal theory	0	0·5000	0	2
	Standard Error	0·0158	0·0079	($\sigma_{\sqrt{\beta_1}}=0\cdot0447$)	0·0447
1000 samples of 5 from triangular population ($\rho=0\cdot5$)	Experiment	0·4653	0·4127	1·3455⁻	3·9442
	Normal theory	0·4517	0·4239	1·0315	3·4191
	Standard Error	0·0134	0·0104	0·1312	0·2096
500 samples of 10 from triangular population ($\rho=0\cdot5$)	Experiment	0·4910	0·2483	0·8882	5·4425
	Normal theory	0·4787	0·2671	0·7431	3·6774
	Standard Error	0·0119	0·0098	0·2085⁺	0·4737
1000 samples of 5 from normal population ($\rho=0\cdot83$ for discrete frequencies and $\rho=0\cdot9$ for continuous)	Experiment	0·7852	0·2306	5·0457	9·2953
	Normal theory $\rho=0\cdot80$	0·7541	0·2691	5·4065	9·7830
	Standard Error	0·0085⁺	0·0126	—	—
	" $\rho=0\cdot83$	0·7873	0·2461	—	—
	" $\rho=0\cdot90$	0·8687	0·1748	13·0290	21·7579
	Standard Error	0·0055⁺	0·0127	—	—
500 samples of 10 from normal population (as for samples of 5)	Experiment	0·8012	0·1433	3·5357	9·4080
	Normal theory $\rho=0\cdot80$	0·7819	0·1461	3·1377	8·0534
	Standard Error	0·0065⁺	0·0087	—	—
	" $\rho=0\cdot83$	0·8135⁺	0·1288	3·72	9·2
	" $\rho=0\cdot90$	0·8887	0·0832	5·7475	13·6667
	Standard Error	0·0037	0·0066	—	—

The standard errors shown are all approximate. They were obtained as follows: The standard errors of β_1 and β_2 were taken from Tables XXXVII and XXXVIII of Karl Pearson's *Tables for Statisticians and Biometricians*, Part I. In the case in which $\beta_1=0$ the approximate formula†

$$\sigma_{\sqrt{\beta_1}}=(\beta_4-6\beta_2+9)^{\frac{1}{2}}/N^{\frac{1}{2}}$$

* The values corresponding to $\rho=0\cdot9$ and $\rho=0\cdot8$ were found directly in *Biometrika*, Vol. XI (1915—17), pp. 399 ff. The values corresponding to $\rho=0\cdot83$ were obtained by interpolation.

† See E. S. Pearson, "The Test of Significance for the Correlation Coefficient," *Journal of the American Statistical Association*, Vol. XXVI (1931), p. 18.

for the standard error of $\sqrt{\beta_1}$ was used. For the other constants the formulae*

$$\sigma_{\bar{r}} = \frac{\sigma}{\sqrt{N}}, \quad \sigma_\sigma = \frac{\sigma}{2}\left(\frac{\beta_2 - 1}{N}\right)^{\frac{1}{2}}$$

were employed. Here N is the number of samples, which in the present cases is either 1000 or 500.

In most instances the deviations of the constants from their expected values in samples from normal populations are not improbable. The worst discrepancy seems to be that of β_2 in the 500 samples of 10 from the triangular population. For the samples from the normal population having high correlation the constants seem to be in substantial agreement with the theoretical values for samples from a continuous normal population having $\rho = 0{\cdot}83$, which is the value obtained from the sampled correlation table without Sheppard's correction for grouping.

II. THE EFFECT OF THE COARSENESS OF GROUPING.

Description of the Populations sampled.

To study the effect of the coarseness of grouping upon the distribution of r two correlation tables were constructed from tables of volumes of the normal surface† for $\rho = 0{\cdot}5$. These are shown in Tables VIII and IX and are designated Populations A and B respectively.

TABLE VIII.

Population A.

								Totals
	0	1	7	32	67	64	30	201
	1	11	74	229	326	212	64	917
	7	74	346	742	727	326	67	2289
	32	229	742	1097	742	229	32	3103
	67	326	727	742	346	74	7	2289
	64	212	326	229	74	11	1	917
	30	64	67	32	7	1	0	201
Totals	201	917	2289	3103	2289	917	201	9917

$\rho = 0{\cdot}4602$, without Sheppard's correction.
$\rho = 0{\cdot}49008$, with Sheppard's correction.

Population A. This is a 7×7 cell table. The class interval is $0{\cdot}80\sigma$, or $0{\cdot}82\sigma$ if Sheppard's correction has been applied. The value of ρ as actually calculated from the table is 0·4602. With Sheppard's correction (applied to the two marginal standard

* See footnote † on p. 396.

† See *Tables for Statisticians and Biometricians*, Part II, Table VIII, pp. 78—135.

deviations involved in ρ), $\rho = 0{\cdot}49008$, essentially the same as in the continuous frequency surface.

TABLE IX.

Population B.

						Totals
	0	9	90	170	68	337
	9	227	1028	940	170	2374
	90	1028	2273	1028	90	4509
	170	940	1028	227	9	2374
	68	170	90	9	0	337
Totals	337	2374	4509	2374	337	9931

$\rho = 0{\cdot}4377$, without Sheppard's correction.
$\rho = 0{\cdot}4924$, with Sheppard's correction.

Population B. This is a 5×5 cell table. The class interval is $1{\cdot}16\sigma$, or $1{\cdot}23\sigma$ if Sheppard's correction has been applied. The value of ρ calculated from the table is $0{\cdot}4377$. With Sheppard's correction, $\rho = 0{\cdot}4924$.

Results of Sampling. Comparison of Constants.

Five hundred samples of 5 pairs each were drawn (by Tippett's numbers) from Population A and also from Population B*. These were combined to form 250 samples of 10 from each of the populations.

In the case of samples of 10 from Population A, the values of r were calculated both without and with Sheppard's correction. In the rather rare instances in which this correction led to a value of r greater than unity the value $r = 1$ was used.

The resulting distributions of r are shown in Table X and are compared with the corresponding theoretical distributions from a continuous population in Tables XI and XII through the values of the means, the standard deviations, and the two betas. There is some indication, supported by the result of the goodness of fit test provided in Tables XIII and XIV below, that the distribution of r for samples from a grouped normal population tends to be the same as that for samples from

* Twenty indeterminate forms occurred in calculating r from samples from Population B. By indeterminate form is meant a sample in which one of the standard deviations used in calculating r is zero (or in which both of them are zero). A typical example that actually occurred is

x	-1	0	0	-2	-1
y	0	0	0	0	0

Each sample like this was discarded and replaced by a new sample.

a continuous population having ρ equal to the value obtained from the grouped population without applying Sheppard's correction. This tendency is also marked in the samples from the normal population having high correlation. (See Part I.)

TABLE X.

Distribution of r in Samples of 5 and of 10 from Populations A and B.

r	Frequencies			
	A_5	B_5	A_{10}	B_{10}
·95 to 1·00	14	15	1	1
·90 „ ·95	26	33	1	4
·85 „ ·90	27	34	4	2
·80 „ ·85	21	28	4	8
·75 „ ·80	33	31	18	15
·70 „ ·75	28	12	14	14
·65 „ ·70	26	19	21	15
·60 „ ·65	38	41	17	20
·55 „ ·60	20	15	23	26
·50 „ ·55	33·5	27	18	20
·45 „ ·50	27·5	21	15	15
·40 „ ·45	27	30	14	17
·35 „ ·40	16	18	18	7
·30 „ ·35	12	17	15	17
·25 „ ·30	13·5	13	12	17
·20 „ ·25	7·5	18	11	7
·15 „ ·20	13	11	7	8
·10 „ ·15	11	10	7	4
·05 „ ·10	19	5	7	4
·00 „ ·05	9·5	19	7	5
·00 „ −·05	10·5	17	2	7·5
−·05 „ −·10	3	0	5	6·5
−·10 „ −·15	4	7	2	2
−·15 „ −·20	7	6	1	3
−·20 „ −·25	8	9	3	2
−·25 „ −·30	5	4	1	2
−·30 „ −·35	3	3	—	1
−·35 „ −·40	4	3	2	1
−·40 „ −·45	6	12	—	2
−·45 „ −·50	2	1	—	—
−·50 „ −·55	6	4	—	1
−·55 „ −·60	3	1	—	—
−·60 „ −·65	2	6	—	—
−·65 „ −·70	3	3	—	—
−·70 „ −·75	2	1	—	—
−·75 „ −·80	3	3	—	—
−·80 „ −·85	2	1	—	—
−·85 „ −·90	2	—	—	—
−·90 „ −·95	2	—	—	—
−·95 „ −1·00	—	2	—	—
Totals	500	500	250	250

A_5 and A_{10} refer to samples of 5 and of 10 respectively from Population A.
B_5 and B_{10} refer to samples of 5 and of 10 respectively from Population B.

TABLE XI.

Constants of Distribution of r for 500 Samples of 5.

Population	$\bar{r}$	σ	β_1	β_2
Experimental:				
A (7×7 cells), $\rho=0{\cdot}4602$, $^*\rho_c=0{\cdot}49008$	0·42305	0·4255⁺	1·1281	3·5818
B (5×5 cells), $\rho=0{\cdot}4377$, $^*\rho_c=0{\cdot}4924$	0·4224	0·4279	0·8340	3·2480
Theoretical:				
Continuous distribution, $\dagger\rho=0{\cdot}5$	0·4517	0·4239	1·0315	3·4191
,, ,, $\rho=0{\cdot}4$	0·3584	0·4528	0·5909	2·8097
,, ,, $\rho=0{\cdot}4602$	0·4143	0·4363	—	—
,, ,, $\rho=0{\cdot}4377$	0·3933	0·4429	—	—

TABLE XII.

Constants of Distribution of r for 250 Samples of 10.

Population	$\bar{r}$	σ	β_1	β_2
Experimental:				
A (7×7 cells), $\rho=0{\cdot}4602$, $^*\rho_c=0{\cdot}49008$	0·4440	0·2662	0·4400	3·0453
B (5×5 cells), $\rho=0{\cdot}4377$, $^*\rho_c=0{\cdot}4924$	0·4431	0·2794	0·5583	3·2786
A (Sheppard's correction applied to each sample)	0·4780	0·2855⁺	0·5289	3·1995⁺
Theoretical:				
Continuous distribution, $\dagger\rho=0{\cdot}5$	0·4787	0·2671	0·7431	3·6774
,, ,, $\rho=0{\cdot}4$	0·3813	0·2917	0·4374	3·1669
,, ,, $\rho=0{\cdot}4602$	0·4398	0·2777	0·606	3·45⁻
,, ,, $\rho=0{\cdot}4377$	0·4179	0·283	0·538	3·33

* ρ_c is the value of the correlation coefficient when Sheppard's correction has been applied to the marginal standard deviations of the correlation table.

† The values of the constants of the continuous distribution are independent of the number of samples, although dependent on the number in the sample.

Fisher's transformation $z=\tanh^{-1}r$, $\zeta=\tanh^{-1}\rho$, used in Part I, was also used here to obtain theoretical frequencies.

For Population A: $\zeta=\tanh^{-1}0{\cdot}4602=0{\cdot}497565^-$, $\bar{z}=0{\cdot}523562$ by formula (6), $\sigma_z=0{\cdot}373324$ by formula (7); the goodness of fit test, worked out in Table XIII, gives $P=0{\cdot}9503$. For Population B: $\zeta=\tanh^{-1}0{\cdot}4377=0{\cdot}469382$, $\bar{z}=0{\cdot}494101$, $\sigma_z=0{\cdot}373515^-$, $P=0{\cdot}502$. (See Table XIV.)

Summary and Conclusions.

Actual samples of 5 and of 10 were drawn from bivariate populations differing greatly from the normal. Population correlations $\rho=0$ and $\rho=0{\cdot}5$ were used. The distributions of r were not essentially different from the theoretical distributions in samples from a normal population which may be considered in fact as providing

TABLE XIII.

Goodness of Fit. Samples of 10 *from Population A. Fisher's Transformation.*

r	$z=\tanh^{-1}r$	$\frac{z-\bar{z}}{\sigma_z}$	Normal area $\frac{1}{2}(1+\alpha)$	Δ	$250\Delta=f$	Observed frequency f_0	$\frac{(f_0-f)^2}{f}$
1·00	∞	∞	1·00000				
·95	1·831 7808	3·5042	·99977	·00023	·06	1	
·90	1·472 2195	2·5411	·99447	530	1·32	1	·01
·85	1·256 1528	1·9623	·97515	1932	4·83	4	
·80	1·098 6123	1·5404	·93827	3688	9·22	4	2·96
·75	·972 9551	1·2038	·88570	5257	13·14	18	1·80
·70	·867 3005*	·9208	·82147	6423	16·06	14	·26
·65	·775 2987	·6743	·74997	7150	17·88	21	·54
·60	·693 1472	·4543	·67508	7489	18·72	17	·16
·55	·618 3813	·2540	·60025	7483	18·71	23	·98
·50	·549 3061	·0690	·52751	7274	18·18	18	·00
·45	·484 7003	− ·1041	·52141	6892*	17·23	15	·29
·40	·423 6489	− ·2676	·60550	6409	16·02	14	·25
·35	·365 4438	− ·4235+	·66403	5853	14·63	18	·71
·30	·309 5196	− ·5733	·71679	5276	13·19	15	·25
·25	·255 4128	− ·7183	·76362	4683	11·71	12	·01
·20	·202 7326	− ·8594	·80483	4121	10·30	11	·05
·15	·151 1404	− ·9976	·84074	3591	8·98	7	·44
·10	·100 3353	−1·1337	·87154	3080	7·70	7	·06
·05	·050 0417	−1·2684	·89760	2606	6·52	7	·04
·00	·000 0000	−1·4024	·91962	2202	5·50	7	·41
−·05	−·050 0417	−1·5365−	·93779	1817	4·54	2	·19
−·10	−·100 3353	−1·6712	·95264	·01485	3·71	5	
−·15						2	
−·20						1	
−·25				·04736	11·84	3	·68
−·30						1	
−·35							
−·40						2	
					250·00	250	10·09

$\bar{z}=\cdot523{,}562$, $\sigma_z=\cdot373{,}324$, $\chi^2=10\cdot09$, $n=19$, $P=0\cdot9503$.

* Note that we have a positive and a negative value of $(z-\bar{z})/\sigma_z$.

good first approximations*. If this is true for samples containing so few items as 5 or 10, it will assuredly hold true for larger samples.

The values of r in actual samples ($n=5$ and $n=10$) from a normal population having high correlation seem to follow the theoretical distribution curve.

* Cf. E. S. Pearson, "Some Notes on sampling Tests with two Variables," *Biometrika*, Vol. xxi^{B} (1929), pp. 337—360. In commenting on certain of his experiments, Pearson says, "These two series of results are of considerable interest and suggest that the normal bivariate surface can be mutilated and distorted to a remarkable degree without affecting the frequency distribution of r in samples as small as 20." (p. 357.)

See G. A. Baker, "The Significance of the Product-Moment Coefficient of Correlation with special reference to the Character of the Marginal Distributions," *Journal of the American Statistical Association*, Vol. xxv (1930), pp. 387—396. See also E. S. Pearson, "The Test of Significance for the Correlation Coefficient," *Journal of the American Statistical Association*, Vol. xxvi (1931), pp. 128—134. Read E. S. Pearson's comments on Baker's results. [The conclusion above seems not wholly consistent with the P's of Tables V and VI. ED.]

TABLE XIV.

Goodness of Fit. Samples of 10 *from Population B. Fisher's Transformation.*

r	$z=\tanh^{-1} r$	$\frac{z-\bar{z}}{\sigma_z}$	Normal area $\frac{1}{2}(1+\alpha)$	Δ	$250\Delta=f$	Observed frequency f_0	$\frac{(f_0-f)^2}{f}$
1·00	∞	∞	1·00000	·00017	·04	1	
·95	1·831 7808	3·5813	·99983	393	·98	4	·66
·90	1·472 2195	2·6187	·99590	1658	4·14	2	
·85	1·256 1528	2·0402	·97932	3211	8·03	8	·01
·80	1·098 6123	1·6184	·94721	4713	11·78	15	·88
·75	·972 9551	1·2820	·90008	5898	14·74	14	·37
·70	·867 3005	·9992	·84110	6689	16·72	15	·18
·65	·775 2987	·7528	·77421	7123	17·81	20	·27
·60	·693 1472	·5329	·70298	7266	18·16	26	3·38
·55	·618 3813	·3327	·63032	7157	17·89	20	·25
·50	·549 3061	·1478	·55875	6880*	17·20	15	·28
·45	·484 7003	− ·0252	·51005	6475	16·19	17	·04
·40	·423 6489	− ·1886	·57480	5994	14·98	7	4·25
·35	·365 4438	− ·3445	·63474	5460	13·65	17	·82
·30	·309 5196	− ·4942	·68934	4925	12·31	17	1·78
·25	·255 4128	− ·6390	·73859	4374	10·94	7	1·42
·20	·202 7326	− ·7801	·78233	3841	9·60	8	·27
·15	·151 1404	− ·9182	·82074	3336	8·34	4	2·26
·10	·100 3353	− 1·0542	·85410	2868	7·17	5	·66
·05	·050 0417	− 1·1889	·88278	2430	6·08	7·5	·33
·00	·000 0000	− 1·3228	·90708	2036	5·09	6·5	·04
− ·05	− ·050 0417	− 1·4568	·92744	1681	4·20	2	
− ·10	− ·100 3353	− 1·5915−	·94425	1371	3·43	3	1·10
− ·15	− ·151 1404	− 1·7275−	·95796	·01099	2·75	2	
− ·20	− ·202 7326	− 1·8656	·96895			2	
− ·25						1	
− ·30						1	
− ·35				·03105	7·76	2	·07
− ·40							
− ·45						1	
− ·50							
					249·98	250	19·32

$\bar{z}=\cdot 494{,}101$, $\sigma_z=\cdot 373{,}515$, $\chi^2=25\cdot 22$, $n=20$, $P=0\cdot 502$.

* Note that we have a positive and a negative value of $(z-\bar{z})/\sigma_z$.

Samples of 5 and of 10 were taken from coarsely grouped normal correlation tables. The value of ρ in these tables was 0·5 if Sheppard's correction was applied, without this correction it was somewhat less. The distribution of r in these samples seems to be essentially the same as the theoretical distribution of r in samples from a continuous population in which the value of ρ is that obtained from the correlation table without applying Sheppard's correction. This might be taken as indicating that Sheppard's correction should be applied to the sample†, although this appears only partially to make the distribution the same as that for samples from a continuous population. It is undoubtedly better to avoid coarse grouping.

† See R. A. Fisher, *Statistical Methods for Research Workers*, p. 152.

BIBLIOGRAPHY.

BAKER, GEORGE A. The Significance of the Product-Moment Coefficient of Correlation with special reference to the Character of the Marginal Distributions. *Journal of the American Statistical Association*, Vol. XXV. (1930), pp. 387—396.

DUNLAP, HILDA FROST. An empirical Determination of the Distribution of Means, Standard Deviations and Correlation Coefficients drawn from Rectangular Populations. *Annals of Mathematical Statistics*, Vol. II. No. 1 (1931), pp. 66—81.

EZEKIEL, MORDECAI. *Methods of Correlation Analysis*. New York, Wiley (1930).

—— The sampling Variability of linear and curvilinear Regressions. *Annals of Mathematical Statistics*, Vol. I. (1930), pp. 275—333.

FISHER, R. A. *Statistical Methods for Research Workers*. Edinburgh and London, Oliver and Boyd, 1st edition, 1925; 2nd edition, 1928; 3rd edition, 1930.

—— Frequency Distribution of the Values of the Correlation Coefficient in samples from an indefinitely large population. *Biometrika*, Vol. X. (1914—15), pp. 507—521.

PEARSON, EGON S. Some Notes on sampling Tests with two Variables. *Biometrika*, Vol XXI[B]. (1929), pp. 337—360.

—— The Test of Significance for the Correlation Coefficient. *Journal of the American Statistical Association*, Vol. XXVI. (1931), pp. 128—134.

SOPER, H. E. On the Probable Error of the Correlation Coefficient to a second Approximation. *Biometrika*, Vol. IX. (1913), pp. 91—115.

SOPER, H. E.; YOUNG, A. W.; CAVE, B. M.; LEE, A.; and PEARSON, K. On the Distribution of the Correlation Coefficient in small Samples. Appendix II to the papers of "Student" and R. A. Fisher. *Biometrika*, Vol. XI. (1915—17), pp. 328—413.

"STUDENT." Probable Error of a Correlation Coefficient. *Biometrika*, Vol. VI. (1908—09), pp. 302—310.

THE PERCENTAGE LIMITS FOR THE DISTRIBUTION OF RANGE IN SAMPLES FROM A NORMAL POPULATION. ($n \leqslant 100$.)

By EGON S. PEARSON, D.Sc.

The Table A given on p. 416 below represents an attempt to summarise in most convenient form for practical use the recent work on the distribution of range in samples from a normal population*. It deals only with the case of samples of 100 or less. The accompanying discussion may be divided into three parts:

(1) The method of computation of Table A. (See p. 416.)

(2) Experimental checks on the adequacy of the approximation involved.

(3) Illustrations of the use of Table A.

In addition to the value of mean and standard deviation, of which the former has already been completely and the latter partially tabled, the present Table gives certain percentage limits, namely the values of the range which will (*a*) not be attained, and (*b*) be exceeded, in 0·5 %, 1 %, 5 %, and 10 % of random samples. The unit is throughout the population standard deviation.

(1) *The Method of Computation of Table A.*

It has been assumed that the sampling distribution of range may be adequately represented by Pearson curves with the appropriate moment-coefficients. That this assumption is not unreasonable will be seen from the experimental results presented below, but since it has had to be made, no very high degree of accuracy is justified in calculating the percentage limits from these curves. Nor indeed is a high degree of accuracy required for practical purposes. The procedure adopted may be summarised as follows:

(*a*) A framework was first obtained by finding the equations of the Type I and Type VI curves, using the appropriate frequency constants (set out in Table VIII of my paper referred to above), for samples of size

$$n = 3,\ 4,\ 6,\ 10,\ 20,\ 60,\ 100.$$

The first four of these curves were made to start at the point, range $= w = 0$, and given the correct mean ($= \overline{w}$), standard deviation ($= \sigma_w$), and β_1. For the curves at $n = 20$, 60 and 100 the start was not fixed and the first four theoretical moment

* L. H. C. Tippett, *Biometrika*, Vol. XVII. pp. 364—387. E. S. Pearson, *Biometrika*, Vol. XVIII. pp. 173—194. "Student," *Biometrika*, Vol. XIX. pp. 151—164. *Tables for Statisticians and Biometricians*, Part II, pp. cx—cxix.

coefficients were used—$\bar{w}$, σ_w, β_1 and β_2. The curves given by "Student"* were all calculated by the second method; since however the distribution of range is abrupt at the lower end when n is small, it seemed probable that a better representation of the true but unknown sampling distribution would be obtained in these cases by making the approximative curve have the correct start. By the time $n = 20$ it was considered of greater importance to use the correct β_2 rather than the correct start.

For the case $n = 10$ the percentage limits were however found both from the fixed start and the 4-moment curve, with the following results.

TABLE I.

	Lower				Upper			
Per cent. Limits	0·5 %	1 %	5 %	10 %	10 %	5 %	1 %	0·5 %
Fixed start Curve	1·35	1·48	1·86	2·09	4·13	4·48	5·15	5·40
4-moment Curve	1·32	1·46	1·86	2·10	4·13	4·47	5·16	5·42

The figures provide some idea of the order of uncertainty involved in the method of approximation used. The addition of a 3rd decimal place in the limits would clearly be meaningless, but the retention of the 2nd decimal appears worth while†.

(b) For each of these framework curves the position of the ordinate at the lower and upper tails cutting off 0·5 %, 1·0 %, 5·0 %, and 10·0 % of the total frequency was found by quadrature and backward interpolation. If w_p represents the range value corresponding to any one of these ordinates, then the quantities

$$l_p = (w_p - \bar{w})/\sigma_w \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

were calculated. For a given per cent. limit, p, the value of l_p will change with n, that is to say with the β_1 and β_2 or shape of the sampling curve. But the change is not very rapid, and it was found possible to interpolate in the framework so as to find with the desired accuracy each of the 8 values of l_p for

$$n = 3,\ 4,\ 5,\ \ldots\ 29,\ 30,\ 35,\ 40,\ \ldots\ 95,\ 100‡.$$

(c) Having calculated the l_p's, it was only necessary to invert the formula (1), and obtain w_p from

$$w_p = \bar{w} + l_p\sigma_w \quad \ldots\ldots\ldots\ldots\ldots\ldots(2).$$

The complete set of values of $\bar{w}$ was given by Tippett, but σ_w had only been computed for $n = 2$, 3, 4, 5, 6, 10, 20, 60 and 100. Three additional values were therefore computed at $n = 30$, 45 and 75 by the same process of cubature as that employed by Tippett, with the following result:

Sample Size	n	30	45	75
Standard Deviation of Range	σ_w	·6927	·6601	·6237

* *Loc. cit.* p. 163.

† The Table on p. 162 of "Student's" paper referred to above gives the limits to the 1st decimal place only.

‡ A graphical method of interpolation was used. A similar process was followed in finding the 1 % and 5 % limits for $\sqrt{\beta_1}$ and β_2; see *Biometrika*, Vol. XXII. p. 247.

TABLE II.

Comparison of Observed and Theoretical Frequency Distributions.

Range (central values)	$n=3$ Observation	$n=3$ Theory	$n=4$ Observation	$n=4$ Theory	Range	$n=10$ Observation	$n=10$ Theory	Range	$n=20$ Observation	$n=20$ Theory	$n=60$ Observation	$n=60$ Theory
·1	5	4·0		0·2	1·1 and less		1·3	1·9 and less		1·6		
·2	16	9·2	1	1·0	1·2	1	1·3	2·0	1	1·5		
·3	12	15·2	3	2·6	1·3	1	2·3	2·1	3	2·6		
·4	24	21·1	2	5·1	1·4	4	3·6	2·2	7	4·1		
·5	18	26·5	7	8·3	1·5	6	5·5	2·3	8	6·3		
·6	32	31·3	9	12·1	1·6	6	8·0	2·4	7	9·1		
·7	41	35·4	16	16·2	1·7	9	11·0	2·5	11	12·7		
·8	34	38·7	18	20·6	1·8	15	14·5	2·6	14	16·9		
·9	38	41·3	19	24·9	1·9	17	18·5	2·7	19	21·7		
1·0	39	43·2	34	29·0	2·0	20	22·8	2·8	29	27·0		
1·1	51	44·5	26	32·8	2·1	30	27·3	2·9	28	32·4		
1·2	29	45·1	37	36·2	2·2	38	31·8	3·0	36	37·8		
1·3	41	45·2	33	39·1	2·3	32	36·2	3·1	30	42·9		1·3
1·4	39	44·8	31	41·4	2·4	41	40·2	3·2	56	47·4		
1·5	45	43·9	49	43·2	2·5	49	43·7	3·3	46	51·0	4	1·1
1·6	54	42·7	50	44·3	2·6	50	46·5	3·4	58	53·7	1	1·9
1·7	51	41·2	47	44·9	2·7	46	48·6	3·5	59	55·3	3	2·8
1·8	34	39·4	48	44·9	2·8	48	49·9	3·6	59	55·7	5	4·1
1·9	37	37·4	47	44·5	2·9	49	50·4	3·7	55	55·1	5	5·7
2·0	38	35·3	46	43·6	3·0	51	50·1	3·8	59	53·5	8	7·4
2·1	32	33·1	45	42·2	3·1	48	49·0	3·9	45	51·0	15	9·2
2·2	28	30·9	43	40·6	3·2	44	47·3	4·0	56	47·8	8	11·1
2·3	25	28·6	28	38·7	3·3	56	45·1	4·1	36	44·1	12	12·7
2·4	24	26·3	46	36·5	3·4	30	42·4	4·2	44	40·1	12	14·1
2·5	32	24·1	41	34·2	3·5	44	39·3	4·3	44	36·0	11	15·2
2·6	23	21·9	33	31·8	3·6	48	36·1	4·4	40	31·8	17	15·8
2·7	25	19·8	27	29·3	3·7	28	32·7	4·5	29	27·8	19	16·0
2·8	14	17·8	24	26·9	3·8	28	29·3	4·6	22	23·9	17	15·8
2·9	21	15·9	26	24·4	3·9	25	26·0	4·7	20	20·4	12	15·2
3·0	16	14·2	22	22·0	4·0	20	22·8	4·8	13	17·2	14	14·3
3·1	15	12·5	22	19·8	4·1	15	19·9	4·9	10	14·3	14	13·2
3·2	11	11·0	8	17·6	4·2	17	17·1	5·0	11	11·8	11	11·9
3·3	10	9·6	15	15·5	4·3	13	14·6	5·1	13	9·7	11	10·6
3·4	5	8·3	19	13·6	4·4	15	12·3	5·2	3	7·8	6	9·2
3·5	6	7·1	8	11·9	4·5	13	10·3	5·3	6	6·3	9	7·9
3·6	3	6·1	16	10·3	4·6	10	8·5	5·4	7	5·0	5	6·7
3·7	3	5·2	11	8·8	4·7	3	7·0	5·5	5	3·9	11	5·6
3·8	6	4·4	9	7·5	4·8	7	5·7	5·6	—	3·1	3	4·6
3·9	2	3·6	7	6·3	4·9	7	4·6	5·7	3	2·4	4	3·7
4·0	5	3·0	8	5·3	5·0	6	3·7	5·8	3	1·8	3	3·0
4·1	2	2·5	4	4·4	5·1	3	3·0	5·9	2	1·4	2	2·3
4·2	2	2·0	1	3·7	5·2	1	2·3	6·0	1	1·1	4	1·8
4·3	1	1·6	2	3·0	5·3	1	1·8	6·1	—	0·8	2	1·4
4·4	3	1·3	3	2·4	5·4	1	1·4	6·2	1	0·6	1	1·1
4·5	2	1·0	2	1·9	5·5	1	1·1	6·3	1	0·5	—	0·8
4·6 and more	6	2·8	7	6·5	5·6 and more	3	3·2	6·4 and more	—	1·2	1	2·5
Totals	1000	1000·0	1000	1000·0		1000	1000·0		1000	1000·0	250	250·0

From this framework the intermediate values of σ_w shown in Table A were readily obtained, with an error which it is believed should not be greater than a unit in the 3rd decimal place. Finally the limits, w_p, were obtained from equation (2) and are given in the Table*. How close these approximations are to the true values we cannot at present tell, but the results described in the following section suggest that they are not seriously in error.

(2) *Experimental Checks on the Adequacy of the Approximation involved.*

Tippett has given three experimental sampling distributions for $n = 5$, 10 and 20, and has fitted theoretical curves to the last two. But he placed little weight on the values of β_1 and β_2 used†; since then improved values have been suggested and it is on these tha. the present Table A is based. As I had available seven further series of samples, use was made of these for a fresh comparison. The series consisted of 1000 random samples of sizes $n = 3$, 4, 5, 7, 10, 15 and 20 drawn from a normal population‡. The seven series were completely independent, but a further series of 250 samples of 60 was obtained by combining together the samples of 15 in groups of 4. The observed and theoretical frequencies are compared in Table II, for $n = 3$, 4, 10, 20 and 60. The result of applying the test for goodness of fit is shown below in Table III. In calculating χ^2 small groups were combined at the tails of the distributions so that none of the theoretical frequency groups contain a frequency of less than 5. The brackets in Table II show the groupings used.

TABLE III.

Sample size	3	4	10	20	60
χ^2	43·00	34·99	26·28	28·10	15·93
n'	39	40	38	35	23
P	·266	·653	·905	·752	·819

For the cases $n = 5$, 7 and 15, for which the theoretical curves of the framework were not available, only the expected and observed frequencies lying outside the eight percentage limits given in Table A are shown (Table IV).

The least satisfactory agreement occurs when $n = 3$, where the curve appears to allow for too few cases at the two extremes, particularly at the lower limit. But apart from this there seems little evidence of any systematic differences, and taken as a whole the results encourage a reasoned confidence in the use of the percentage limits in Table A.

* For the case $n=2$, the distribution of w is the half of a normal curve whose standard deviation (if complete) is $\sqrt{2}$. The limits were found from Sheppard's Tables.

† *Loc. cit.* p. 378. Reasons for modifying Tippett's values of β_1 and β_2 were discussed by the present writer in the paper referred to above.

‡ The samples of 3 were provided by Dr J. F. Tocher and those of 20 by Professor T. Hojo; the remainder were drawn for me by Mr A. E. Stone. I take this opportunity of thanking them all heartily for their assistance. In all cases the sampling was carried out with the aid of Tippett's Random Sampling Numbers (*Tracts for Computers*, No. xv). The group breadth was $\frac{1}{10}$ the population standard deviation.

TABLE IV.

Frequencies outside °/₀ Limits ($n = 5$, 7 *and* 15).

		Lower Limits				Upper Limits			
		0·5 °/₀	1 °/₀	5 °/₀	10 °/₀	10 °/₀	5 °/₀	1 °/₀	0·5 °/₀
Expected ...		5	10	50	100	100	50	10	5
Observed	$n = 5$	3	10	55	97	82	42	12	5
	$n = 7$	6	12	45	102	88	53	14	10
	$n = 15$	6	12	59	100	100	51	12	6

(3) *Illustrations of the Use of Table A.*

In applying statistical analysis to test the probability of a given hypothesis, there will often be more than one method of procedure and more than one criterion which may be used. Thus in testing on a sample or samples some hypothesis concerning variation in the sampled population, we may use among possible criteria either the standard deviation or the range. In so far as we know the sampling distribution in both cases, either criterion is of equal value in controlling the risk of rejecting the hypothesis tested when it is true. But in general when dealing with normally distributed variables, tests on the standard deviation will be more efficient in preventing the acceptance of a hypothesis which is false, than those based upon the range. It must also be remembered that the theory assumes random sampling from a homogeneous population, and a single anomalous individual is more likely to throw out a result based upon range than one based upon standard deviation. In certain tests however the range criteria are at less disadvantage than in others, and because of their simplicity in application, if employed with judgment, they will often prove to be extremely useful tools.

Example 1. In order to determine whether a given "lot" of a certain material is up to specification, it may be necessary to consider not only the average value of some character measured on each article, but also its standard deviation, σ. If the lot be large, perhaps composed of several hundred articles, it is a common commercial practice to estimate its nature by sampling. Suppose that it is wished to fix a rejection limit for the variation permissible in the sample in such a way that we are unlikely to reject a lot for which $\sigma \leqslant a$, and unlikely to accept it if $\sigma \geqslant ka$, where a and k (> 1) are to be given some appropriate values depending on the quality of the material which we are prepared to accept. Let us define "unlikely" as corresponding to a 1 in 100 chance. What size of sample will then be necessary to ensure this result if we use (a) the sample range, and (b) the sample standard deviation, to provide the estimate of variation?*

* We shall suppose that evidence is available that the character is distributed approximately normally. Further that the size of the sample is small compared to the size of the lot, so that the sample may be considered as drawn from an "infinite population."

(*a*) *Using Range.* Let $l(n, \cdot 01)$ be the lower and $l(n, \cdot 99)$ the upper 1 % limit obtained from Table A. Our rule will be to reject the lot when the sample range, w, is $> w_0$. To determine w_0 we must find n so that

$$l(n, \cdot 99) \times a = w_0 = l(n, \cdot 01) \times ka \quad \text{.......................(3)}.$$

The first equality will result in the long run in our rejecting a lot for which $\sigma \leqslant a$ at most 1 time in 100; the second, in our accepting a lot for which $\sigma \geqslant ka$ at most 1 time in 100. Suppose now that $k = 2$. An examination of the Table shows that for $n = 40$, $l(n, \cdot 99)/l(n, \cdot 01) = 6\cdot 09/2\cdot 97 = 2\cdot 05$; and for $n = 45$, the ratio is $6\cdot 16/3\cdot 09 = 1\cdot 99$. The desired size of sample is therefore about 44, and $w_0 = 6\cdot 15a$.

(*b*) *Using Standard Deviation.* Let s be the sample standard deviation, and s_0 the limiting value. The upper and lower 1 % limits for s may be found from the tables of the χ^2 integral, where $\chi^2 = ns^2/\sigma^2$, and in Elderton's notation $n' = n$*. Using a similar notation to that of the range problem above, we must make s_0 satisfy the relation

$$\chi^2(n, \cdot 99) \times \frac{a^2}{n} = s_0^2 = \chi^2(n, \cdot 01) \times \frac{k^2 a^2}{n} \quad \text{.....................(4)}.$$

Again taking $k = 2$ for purposes of illustration, it will be found that for $n = 24$, $\chi^2(n, \cdot 99)/\chi^2(n, \cdot 01) = 4\cdot 08$ and for $n = 25$ the ratio is $3\cdot 96$. Consequently the condition will be satisfied for a sample of 25, and $s_0^2 = 1\cdot 72a^2$. Method (*b*) has therefore a clear advantage over method (*a*), for the gain in time in computation following the use of range would hardly balance the loss involved in measuring 20 additional articles.

Example 2. The use of range is however of greater value if a sample is divided into small groups. Suppose that a sample of N is broken up in a random manner into m sub-samples each containing n observations, so that $N = mn$. Let $w_1, w_2, \dots w_m$ be the observed ranges of these sub-samples, and

$$\tilde{w} = (w_1 + w_2 + \dots + w_m)/m \quad \text{..........................(5)}$$

be their mean value. We know that the expected values of the mean range and standard deviation of range in repeated samples of n are

$$\overline{w} = a_n \sigma, \quad \sigma_w = b_n \sigma \quad \text{............................(6)},$$

where a_n and b_n are given in the second and third columns of Table A. It follows that we may use as an estimate of the population standard deviation, σ,

$$\sigma_2 = \tilde{w}/a_n \quad \text{.......................................(7)},$$

and that this estimate will have a standard error (S.E.)

$$\text{S.E. of } \sigma_2 = \frac{1}{a_n} (\text{S.E. of } \tilde{w}) = \frac{b_n}{a_n} \frac{\sigma}{\sqrt{m}} \quad \text{.......................(8)}.$$

* The 1 % limits may be obtained directly without interpolation from the χ^2 tables in R. A. Fisher's *Statistical Methods for Research Workers*; the n of these tables is the number of degrees of freedom = sample size − 1.

Let us compare the reliability of this estimate with that obtained from the standard deviation, s, of the whole $N = mn$ observations*. In this case we know that if $\bar{s}$ is the mean and σ_s the standard error of s in repeated samples, then

$$\bar{s} = c_N \sigma, \quad \sigma_s = d_N \sigma \quad \text{..........................(9)},$$

where the values of c_N and d_N for small samples have been tabled for the case of a normal population†, and tend rapidly to 1 and $1/\sqrt{2N}$, respectively, as N increases. Thus we may use as an estimate of σ

$$\sigma_1 = s/c_N \quad \text{..................................(10)},$$

which will have as its standard error

$$\text{S.E. of } \sigma_1 = \frac{d_N}{c_N}\sigma = \theta_N \frac{\sigma}{\sqrt{2N}} \quad \text{......................(11)},$$

where $\theta_N \to 1$ as $N \to \infty$.

A comparison of the reliability of these two methods of estimating σ is obtained from the ratio

$$\frac{\text{Standard Error of } \sigma_2}{\text{Standard Error of } \sigma_1} = \frac{1}{\theta_N}\frac{b_n\sqrt{2n}}{a_n} = \frac{\phi_n}{\theta_N} \quad \text{.................(12)}.$$

The following are a few values of θ_N:

N	5	10	20	30	40	50	100
θ_N	1·148	1·068	1·033	1·021	1·016	1·013	1·006

By taking $\theta_N = 1$ we shall slightly underestimate the reliability of σ_2 (compared to σ_1), but the correction can be made if desired. A series of values of $\phi_n = \sqrt{2n}\, b_n/a_n$ is given in Table V. It is seen that the range method of estimation may be used to best advantage by breaking up the observations into equal random sub-samples of from 6 to 10 individuals. If this is done, the standard error of $\sigma_2 =$ Mean range$/a_n$ will be approximately 1·15 times the standard error of $\sigma_1 = s$‡.

TABLE V.

n	ϕ_n	n	ϕ_n	n	ϕ_n
2	1·511	12	1·169	30	1·314
3	1·286	13	1·176	35	1·352
4	1·209	14	1·183	40	1·387
5	1·175	15	1·191	45	1·418
6	1·159	16	1·200	50	1·449
7	1·153	17	1·208	60	1·509
8	1·152	18	1·216	70	1·563
9	1·154	19	1·225	80	1·613
10	1·158	20	1·234	90	1·662
11	1·163	25	1·275	100	1·706

* Nothing would be gained when using s by dividing the N observations into groups.

† *Biometrika*, Vol. x. p. 529. *Tables for Statisticians and Biometricians*, Part II, Table XVII.

‡ This is omitting the correcting factors c_N and θ_N.

The table may also be used to compare the reliabilities of different forms of estimate of σ made from range. For example, N observations may be used

(*a*) as m groups of n,

$$\text{Estimate,} \quad \sigma_2' = \frac{\tilde{w}}{a_n}; \quad \text{S.E.} = \frac{b_n}{a_n}\frac{\sigma}{\sqrt{m}} = \phi_n \frac{\sigma}{\sqrt{2N}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (13),$$

(*b*) as one group, so that $m = 1$,

$$\text{Estimate,} \quad \sigma_2'' = \frac{w}{a_N}; \quad \text{S.E.} = \frac{b_N}{a_N}\sigma = \phi_N \frac{\sigma}{\sqrt{2N}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (14).$$

The ratio of the two standard errors of estimate is therefore ϕ_n/ϕ_N. For example, the advantage of breaking a sample of 100 into 10 samples of 10 is clear, for $\phi_{10}/\phi_{100} = 1{\cdot}158/1{\cdot}706 = 0{\cdot}679$, or by using σ_2' rather than σ_2'' we obtain a reduction in standard error of about 32 %.

There is also another advantage. The sampling distribution of range is asymmetrical but approaches most nearly to the normal when n lies between 6 and 10*; if β_1 and β_2 refer to the sampling distribution of w, then the coefficients for the sampling distribution of $\tilde{w}$, the mean range in m samples (and therefore the coefficients for $\sigma_2 = \tilde{w}/a_n$), will be $B_1 = \beta_1/m$ and $B_2 = 3 + (\beta_2 - 3)/m$. Consequently the estimate σ_2 will not only have the lowest standard error when n has a value of 6 to 10, but will also be more nearly normally distributed than if n were larger.

Example 3. Reliability in estimation is closely associated with efficiency in discrimination. The following figures represent 40 random variates obtained from sampling a normal distribution with mean $= 51$ and standard deviation $= 10$ units. Could we be sure that they were not drawn from a distribution in which $\sigma = 5$ units?

48, 54, 41, 53, 49; 51, 44, 34, 62, 54; 59, 39, 45, 57, 49;
44, 50, 57, 37, 50; 62, 57, 51, 59, 54; 35, 49, 36, 63, 46;
53, 41, 47, 39, 59; 54, 44, 61, 63, 44.

If the sample is treated as a whole, the lowest and highest variates are found to be 34 and 63, giving a range of 29. For a sample of $n = 40$ from a distribution with $\sigma = 5$, Table A shows that the upper 1 % limit of range is $6{\cdot}09 \times 5 = 30{\cdot}45$. The observed range is slightly less than this, so that the sample would be judged exceptional but not clearly impossible if σ were equal to 5. Suppose now the numbers are broken up into 8 consecutive groups of 5 as shown by the semi-colons; the 8 ranges are now 13, 28, 20, 20, 11, 28, 20, 19. The upper 0·5 % limit for samples of 5 from a distribution with $\sigma = 5$ is now $4{\cdot}85 \times 5 = 24{\cdot}25$, and 2 out of the 8 ranges exceed this value. This alone suggests that the hypothesis, $\sigma = 5$, is unlikely, but we may obtain more convincing evidence by comparing the mean range for the 8 samples, $\tilde{w} = 19{\cdot}87$, with the expected mean and its standard error for $\sigma = 5$.

We have
$$\begin{cases} \overline{w} = a_n \times \sigma = 2{\cdot}3259 \times 5 = 11{\cdot}63, \\ \sigma_{\tilde{w}} = \sigma_w/\sqrt{m} = b_n\sigma/\sqrt{m} = {\cdot}8641 \times 5/\sqrt{8} = 1{\cdot}53. \end{cases}$$

Since $(\tilde{w} - \overline{w})/\sigma_{\tilde{w}} = 5{\cdot}5$, there can now be little doubt whatever that the standard deviation in the sampled population must have been greater than 5 units.

* *Biometrika*, Vol. XVIII. p. 191. *Tables for Statisticians and Biometricians*, Part II, p. cxvii.

Example 4. The figures given in Table VI represent the breaking-strength under tension in lbs. of small briquettes of cement-mortar*. Mixings of the material were carried out on each of 10 different days, and 6 briquettes formed from each mixing. The problem is to determine whether these 10 groups (or samples) of 6 differ significantly either in mean strength or in variability. Other investigations have suggested that the variation in strength within a homogeneous sample is near enough to normal to justify the use of normal theory tests.

TABLE VI.

Breaking-Strength of Cement-Mortar Briquettes.

Group Number ...	1	2	3	4	5	6	7	8	9	10
Values of breaking-strength in lbs.	390	578	530	623	596	345	488	625	722	800
	380	500	540	532	528	322	550	625	727	798
	445	470	470	547	540	312	530	610	690	750
	360	520	560	600	562	358	568	615	700	724
	375	530	460	600	590	375	420	610	705	720
	350	450	470	594	530	310	530	600	700	726
Mean	383·3	508·0	505·0	582·7	557·7	337·0	514·3	614·2	707·3	753·0
Variance	930·6	1733·3	1558·3	1032·6	748·6	588·0	2372·6	78·5	169·2	1150·3
Standard Deviation	30·5	41·6	39·5	32·1	27·4	24·2	48·7	8·9	13·0	33·9
Range	95	128	100	91	68	65	148	25	37	80

In the first place there are two hypotheses to be tested: H_1 that the group variances, and H_2 that the group means do not differ significantly. Let us apply to the problem in turn statistical tests of increasing refinement.

Denote by $\bar{y}_t$, s_t and w_t the mean, standard deviation and range of the tth group; by k the number of groups $(=10)$; by N the total number of observations $(=60)$; by n_t the number in the tth group $(=6)$; and by $\bar{y}_0$ and s_0 the mean and standard deviation of the N observations.

(*a*) *Assume hypothesis H_1 to be true and test H_2. Crude Method using Range.* The mean of the 10 values of w is found to be $\tilde{w} = 83{\cdot}7$ lbs. But for repeated samples of 6 we see from Table A that $\overline{w} = 2{\cdot}53441 \times \sigma$; consequently we may obtain a rough estimate of the assumed common group standard deviation, namely

$$\sigma_2 = 83{\cdot}7/2{\cdot}53441 = 33{\cdot}0 \text{ lbs.} \quad \ldots\ldots\ldots\ldots\ldots\ldots(15).$$

The lowest of the ten group-means is the 6th (337·0 lbs.), and the highest is the 10th (753·0 lbs.); this gives a range of 416·0 lbs. If the means differed only through chance fluctuations, they would vary with a standard error of $\sigma/\sqrt{6}$, which using the estimate $\sigma_2 = 33{\cdot}0$ lbs. becomes 13·5 lbs. The observed range among the ten means

* I am indebted to Mr B. H. Wilson of the Building Research Station for permission to use these data. The cement-mortar used on different days was obtained from different sources so that a difference in mean strength was to be expected.

is $416 \cdot 0/13 \cdot 5 = 30 \cdot 8$ times this standard error. It is almost inconceivable that such a ratio has occurred through chance, and we may therefore conclude that hypothesis H_2 cannot be true, or that the mean strength differs very significantly from group to group*.

It should be noted that when lack of homogeneity has been established either by the range test or otherwise, we may next question whether this is due to the presence in the series of one or more anomalous individual (or group) values. For this purpose Irwin's tables of the probability integral of the distances between the 1st and 2nd, and between the 2nd and 3rd, individuals in samples from a normal population will be of assistance†.

(b) *Assume H_1 true, and test H_2. Exact method.* In this case we avoid the use of range and obtain an estimate of σ^2 from the group variances s_t^2, namely

$$\sigma_1^2 = \sum_{t=1}^{k} (n_t s_t^2)/(N-k) = 1243 \cdot 43, \text{ or } \sigma_1 = 35 \cdot 26 \text{ lbs.} \quad \text{.........(16).}$$

Actually we may avoid the labour of calculating the separate values of s_t^2 by using the identity

$$\sum_{t=1}^{k} (n_t s_t^2) = N s_0^2 - \sum_{t=1}^{k} n_t (\bar{y}_t - \bar{y}_0)^2 \quad \text{....................(17).}$$

It is found that

$$\eta_{y_t}^2 = \sum_{t=1}^{k} n_t (\bar{y}_t - \bar{y}_0)^2/(N s_0^2) = \cdot 9351 \quad \text{....................(18)}$$

and is clearly significant, again showing that hypothesis H_2 is untenable‡.

(c) *To test H_1. Method using Range.* We now wish to determine whether the variation of breaking-strength within each group of 6 changes from group to group more than might be expected through chance. Using the estimate, $\sigma_2 = 33 \cdot 0$ lbs., based upon the mean range we may ask whether the ten group-ranges given at the bottom of Table VI differ significantly. The position can be studied in the upper diagram of Fig. 1, which shows the ten values of range represented by black circles, and also the lower and upper percentage limits obtained by multiplying by $33 \cdot 0$ the figures taken from the row $n = 6$ of Table A. The small figures above the circles refer to the corresponding group numbers of Table VI. It will be seen that 3 ranges out of 10 lie beyond the two 5 % limits; the expectation is 1 out of 10.

The standard deviation of range for samples of 6 is $\cdot 8480 \times \sigma$; using σ_2 as the estimate of σ, we find

$$\sigma_w = 33 \cdot 0 \times \cdot 8480 = 28 \cdot 0 \text{ lbs.} \quad \text{.......................(19).}$$

* This result is of course obvious from a mere inspection of the figures, but the example illustrates the method of attack. Table A shows us that the range among ten means should only exceed $5 \cdot 40 \times$ standard error through chance on 5 occasions in 1000; the experiment gives a factor of 30·8!

† *Biometrika*, Vol. XVII. pp. 238—250. *Tables for Statisticians and Biometricians*, Part II, pp. cv—cx and Tables XIX and XX.

‡ For $N=60$, $k=10$ we should expect in samples were H_2 true, $\bar{\eta}^2 = \cdot 1525$, $\sigma_{\eta^2} = \cdot 0651$. (See Woo's Tables, *Tables for Statisticians and Biometricians*, Part II, p. 17.)

The observed standard deviation of the 10 ranges is larger than this, namely 35·7 lbs.

These results suggest that a more critical analysis is desirable as the variation in group range may perhaps be significant.

(*d*) *Improvement on* (*c*). If test (*b*) has been used in examining the means, the estimate $\sigma_1 = 35{\cdot}26$ lbs. is preferable to $\sigma_2 = 33{\cdot}0$ lbs., and should be used in (*c*). In the present instance, however, the change will scarcely affect the position.

(*e*) *To test* H_1. *Method using Variance.* If it is decided to calculate the individual group variances, s_t^2, we may form a diagram showing the position of the 10 values with regard to the percentage limits in exactly the same manner as for the range. This is shown in the lower half of Fig. 1. For samples of n from a normal distribution, s^2 is distributed according to the law

$$y = y_0 \left(\frac{s^2}{\sigma^2}\right)^{\frac{n-3}{2}} e^{-\frac{ns^2}{2\sigma^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (20).$$

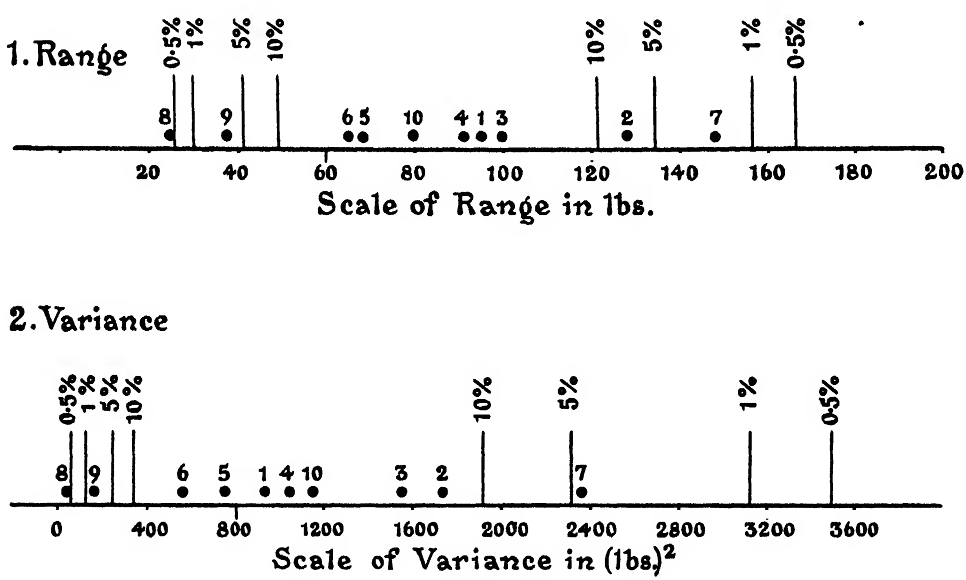


Fig. 1.

The limits may be found either from the *Tables of the Incomplete Gamma Function*, or from those of χ^2 by writing $s^2 = \chi^2\sigma^2/n$. A comparison of the two diagrams shows the small difference in the relative position of the sample points with regard to their scales. Range and variance are in fact highly correlated in small samples, and in the present case the analysis of range, which is far the more rapidly carried out, is probably as useful for the purpose as the analysis of variance.

Both tests as described provide a picture which is of value in forming a judgment on the situation, but neither can lead to an exact measure of probability. For in

each case we must substitute for the unknown σ an estimate (whether σ_1 or σ_2) depending on the individual samples. This disadvantage is overcome in the following test (*f*).

(*f*) *To test* H_1. *Use of the Criterion of Likelihood.* In a paper recently published elsewhere*, Dr J. Neyman and the writer have discussed a test developed from the principle of likelihood. It will perhaps be of interest to conclude by stating briefly the result which it leads to in the present problem. The criterion suggested may be defined as

$$L = \lambda_{H_1}^{\frac{2}{N}} = \sqrt[N]{\prod_{t=1}^{k} (s_t^2)^{n_t}} \Big/ s_a^2 \quad \text{.......................(21)},$$

where

$$s_a^2 = \sum_{t=1}^{n} (n_t s_t^2)/N \quad \text{.............................(22)}.$$

L is therefore the ratio of the weighted geometric mean to the weighted arithmetic mean of the group variances, and is independent of the unknown σ. When the variation is normal, the moments of the sampling distribution of L (if hypothesis H_1 be true) have been found. In the simple case in which the groups contain the same number of individuals, i.e. when $n_1 = n_2 = \ldots = n_k = N/k = n$, say, the pth moment coefficient of L about zero is

$$\mu_p' = k^p \frac{\left\{\Gamma\left(\frac{n-1}{2} + \frac{p}{k}\right)\right\}^k}{\left\{\Gamma\left(\frac{n-1}{2}\right)\right\}^k} \frac{\Gamma\left(\frac{N-k}{2}\right)}{\Gamma\left(\frac{N-k}{2} + p\right)} \quad \text{.(23)}.$$

Reasons are given in the paper referred to for believing that the distribution of L may be represented approximately in many cases by a Type I curve of form

$$y = y_0 L^{m_1 - 1} (1 - L)^{m_2 - 1} \quad \text{.........................(24)}$$

with the correct mean and standard deviation. That is to say, m_1 and m_2 are to be determined from the first two moment coefficients in (23) as follows

$$\left.\begin{aligned} m_1 &= \mu_1' (\mu_1' - \mu_2')/(\mu_2' - \mu_1'^2) \\ m_2 &= (1 - \mu_1')(\mu_1' - \mu_2')/(\mu_2' - \mu_1'^2) \end{aligned}\right\} \quad \text{..................(25)}.$$

The hypothesis H_1 becomes less and less likely as $L \to 0$, and the chance of obtaining $L \leqslant L_0$ may be found by any method giving the values of the Incomplete B-Function.

For the present example, in which $N = 60$, $k = 10$, $n = 6$, it is found from (21) that

$$L = \cdot 702 \quad \text{...................................(26)}.$$

Further $m_1 = 21{\cdot}82$ and $m_2 = 4{\cdot}54$.

We may now apply R. A. Fisher's z-transformation to (24), and write

$$L = \frac{n_2}{n_2 + n_1 e^{2z}} \quad \text{..(27)},$$

$$n_1 = 2m_2 = 9{\cdot}08, \qquad n_2 = 2m_1 = 43{\cdot}64 \quad \text{..................(28)}.$$

* "On the Problem of *k* Samples." *Bulletin de l'Académie Polonaise des Sciences et des Lettres.* Série A. Sciences Mathématiques, 1931, pp. 460—481.

TABLE A.

Percentage Limits for the Distribution of Range in Samples from a Normal Population.

Size of Sample	Mean	Standard Deviation	Lower Limits 0·5 %	Lower Limits 1 %	Lower Limits 5 %	Lower Limits 10 %	Upper Limits 10 %	Upper Limits 5 %	Upper Limits 1 %	Upper Limits 0·5 %	Size of Sample
2	1·12838	·8525	·01	·02	·09	·18	2·33	2·77	3·64	3·97	2
3	1·69257	·8884	·17	·22	·45	·63	2·92	3·34	4·10	4·36	3
4	2·05875	·8798	·38	·47	·77	·98	3·26	3·65	4·38	4·65	4
5	2·32593	·8641	·59	·70	1·04	1·26	3·49	3·87	4·59	4·85	5
6	2·53441	·8480	·78	·89	1·26	1·49	3·67	4·04	4·74	5·00	6
7	2·70436	·833	·95	1·07	1·44	1·68	3·82	4·18	4·87	5·13	7
8	2·84720	·820	1·10	1·22	1·60	1·83	3·94	4·29	4·98	5·23	8
9	2·97003	·808	1·23	1·36	1·74	1·97	4·04	4·39	5·07	5·32	9
10	3·07751	·797	1·35	1·48	1·86	2·09	4·13	4·48	5·15	5·40	10
11	3·17287	·787	1·47	1·59	1·97	2·20	4·21	4·55	5·22	5·47	11
12	3·25846	·778	1·57	1·69	2·07	2·30	4·28	4·62	5·28	5·53	12
13	3·33598	·770	1·66	1·78	2·16	2·39	4·35	4·68	5·34	5·59	13
14	3·40676	·762	1·74	1·86	2·24	2·47	4·41	4·74	5·39	5·64	14
15	3·47183	·755	1·82	1·94	2·32	2·54	4·47	4·79	5·44	5·69	15
16	3·53198	·749	1·89	2·01	2·39	2·61	4·52	4·84	5·49	5·73	16
17	3·58788	·743	1·95	2·08	2·46	2·67	4·57	4·89	5·53	5·77	17
18	3·64006	·738	2·01	2·14	2·52	2·73	4·61	4·93	5·57	5·81	18
19	3·68896	·733	2·07	2·20	2·57	2·79	4·65	4·97	5·61	5·85	19
20	3·73495	·729	2·13	2·25	2·62	2·84	4·69	5·01	5·64	5·89	20
21	3·77834	·724	2·18	2·31	2·67	2·89	4·73	5·05	5·68	5·92	21
22	3·81938	·720	2·23	2·36	2·72	2·93	4·77	5·08	5·71	5·95	22
23	3·85832	·716	2·28	2·40	2·77	2·98	4·80	5·11	5·74	5·98	23
24	3·89535	·712	2·33	2·45	2·81	3·02	4·83	5·14	5·76	6·00	24
25	3·93063	·709	2·37	2·49	2·85	3·06	4·86	5·17	5·79	6·03	25
26	3·96432	·705	2·41	2·53	2·89	3·10	4·89	5·20	5·82	6·05	26
27	3·99654	·702	2·45	2·57	2·93	3·13	4·92	5·23	5·84	6·08	27
28	4·02741	·699	2·49	2·61	2·96	3·17	4·95	5·25	5·87	6·10	28
29	4·05704	·696	2·52	2·65	3·00	3·20	4·97	5·28	5·89	6·12	29
30	4·08552	·693	2·56	2·69	3·04	3·24	5·00	5·30	5·91	6·15	30
35	4·21322	·681	2·72	2·84	3·18	3·38	5·11	5·41	6·01	6·24	35
40	4·32156	·670	2·85	2·97	3·31	3·50	5·20	5·50	6·09	6·32	40
45	4·41544	·660	2·97	3·09	3·42	3·61	5·28	5·57	6·16	6·39	45
50	4·49815	·652	3·07	3·19	3·51	3·70	5·35	5·64	6·23	6·45	50
55	4·57197	·645	3·17	3·28	3·59	3·78	5·42	5·70	6·29	6·51	55
60	4·63856	·639	3·25	3·36	3·67	3·86	5·48	5·76	6·34	6·56	60
65	4·69916	·633	3·32	3·43	3·74	3·93	5·53	5·81	6·38	6·61	65
70	4·75472	·628	3·39	3·50	3·81	3·99	5·58	5·86	6·43	6·65	70
75	4·80598	·624	3·45	3·56	3·87	4·05	5·63	5·91	6·47	6·69	75
80	4·85355	·619	3·51	3·62	3·92	4·10	5·67	5·95	6·50	6·72	80
85	4·89789	·615	3·57	3·67	3·97	4·15	5·71	5·98	6·54	6·76	85
90	4·93940	·612	3·62	3·72	4·02	4·19	5·75	6·02	6·57	6·79	90
95	4·97841	·608	3·67	3·77	4·06	4·24	5·78	6·05	6·60	6·82	95
100	5·01519	·605	3·71	3·81	4·11	4·28	5·81	6·08	6·63	6·85	100

The 5% and 1% limits for z may then be found by interpolating in his tables* with the following results:

$$\left.\begin{array}{lll} 5\% \text{ point} & z = \cdot 371, & L = \cdot 696 \\ 1\% \text{ point} & z = \cdot 519, & L = \cdot 630 \end{array}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots (29).$$

The observed value, $L = \cdot 702$, lies close to the 5% point; or differences in group variation as large or larger than those observed might be expected to arise through chance in about 1 experiment in 20. We can hardly therefore feel confident that they are significant, and find no reason to be dissatisfied by the cruder picture provided by the range test.

In conclusion I would like to acknowledge much friendly assistance received in preparing this paper. Table A is built essentially on the earlier work of Mr L. H. C. Tippett, who has also lent me certain unpublished material which was of much value in its construction. "Student" too has given me many helpful suggestions, and the work was undertaken in the first instance with the knowledge that both Tippett and he had proved the utility of the range criterion in certain fields of practical application.

In addition I am most grateful to Mr M. R. El-Shanawany for a large part of the lengthy computation of the framework distributions for Table A; to Dr L. J. Comrie for supervising the computation of σ_w for the cases $n = 45$ and 75; and to "Mathetes" for the loan of the working sheets on which the theoretical distributions contained in "Student's" 1927 paper on Routine Analysis were based.

* *Statistical Methods for Research Workers*, pp. 212—215.

THE CONVERSE OF SPEARMAN'S TWO-FACTOR THEOREM.

By BURTON H. CAMP, Wesleyan University, Middletown, Conn., U.S.A.

1. Introduction. There have been several attempts to prove, or to disprove, the converse of Spearman's two-factor theorem *. As shown by Irwin, these various methods of proof result ultimately in the same expression for the so-called general factor. Although the several proofs are necessarily alike in many respects, the different authors appear to have different pictures in mind at the background of their analytical demonstrations. The same may be said of the proof presented in this paper. In addition, I have inserted a certain necessary but hitherto neglected † hypothesis, have investigated the possibility of the use of other than linear functions as the basis of the formation of the general factor function, have given a numerical example in which this factor is not unique, and have discussed more fully the important additional question raised by Piaggio as to whether this factor is "almost" unique. Let us begin with the neglected hypothesis.

2. The Net Correlations. It will first be shown that Spearman's proof does not hold in a specific case. The reason it fails is that he has produced a certain determinant as a definition of his general factor without proving that there exists a frequency distribution for which such a definition would be possible. In presenting this example I shall use his notation, to which the reader may wish to refer, unless he chooses to read the later sections of this paper first. My notation, and Spearman's, will be explained in the sections following.

Let r_{ab}, r_{ac}, etc. be as in the following table:

	a	b	c	d
a	—	·8	·5	·35
b	·8	—	·7	·49
c	·5	·7	—	·306
d	·35	·49	·306	—

Evidently, this set forms a perfect hierarchy, thus satisfying Spearman's only hypothesis. He now defines $r_{a\eta}$ so that

$$r_{a\eta}=\frac{1}{\mu_a}=\sqrt{\lambda_{aq}r_{aq}},\quad \lambda_{aq}=r_{ak}/r_{kq},$$

* C. Spearman, *The Abilities of Man* (1927), Macmillan Company, Appendix, pp. ii—vii. E. B. Wilson, *Proceedings of the National Academy of Sciences* (U.S.A), Vol. XIV. (1928), pp. 283—291. H. T. H. Piaggio, *Nature*, January 10, 1931, p. 56. J. O. Irwin, *British Journal of Psychology*, Vol. XXII. (1932), pp. 359—363. Cf. also J. C. M. Garnett, same volume, pp. 364—372.

† Except by Wilson, whose methods are quite different.

where q, k and a are chosen arbitrarily from the letters $a, b, c, \ldots$, except that the same letter must not be chosen twice. In particular, now, let $a=b$, $k=c$, $q=a$. Then

$$r_{b\eta} = \sqrt{\frac{r_{bc} r_{ba}}{r_{ca}}} = \sqrt{\frac{(\cdot 7)(\cdot 8)}{\cdot 5}},$$

which is impossible. The tacit assumption is made, therefore, that

$$r_{ac} - r_{bc} r_{ba} \geqslant 0,$$

and we are led to infer that this inequality should be predicated for all permitted choices of the subscripts, or, to put it another way, that none of the net correlations $r_{ca \cdot b}$ should be permitted to be negative. In fact, this condition is really necessary as will be proved later. On the other hand, I presume that it is true that it would almost certainly be satisfied automatically in the kind of practical problems to which the theory is currently applied, and so it may be that it is only from the logician's standpoint that it is necessary to insist on it.

3. The General Theory. Consider (for definiteness) the aptitudes $a_{\alpha,\beta}$ of $\beta = 1, 2, \ldots, N$ individuals in $\alpha = 1, 2, \ldots, n$ studies. We may always write

(*a*) $$a_{\alpha,\beta} = C_\alpha g_\beta + s_{\alpha,\beta},$$

where C, g and s are functions, to be determined, of their subscripts. (I now depart from Spearman's notation. His f, η, δ and a are the same, respectively, as my C, g, s, and a_1.) The variable a may be supposed, without loss of generality, to have a mean equal to zero, and a σ equal to unity, for each fixed value of α, as β ranges from 1 to N.

(*b*) Let the correlation r_{ij} between a_i and a_j, and the net correlation $r_{ij \cdot k}$ be positive or zero for every set of mutually different values of i, j and k.

(*c*) Let the totality of r_{ij}'s constitute a hierarchy, i.e. let parallel arrays of the following determinant be proportional, if the series of 1's which constitutes the principal diagonal be omitted from consideration:

$$\begin{vmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2n} \\ & & & & \\ & & & & \\ r_{n1} & r_{n2} & r_{n3} & \cdots & r_{nn} \end{vmatrix}$$

For convenience, we shall also suppose this determinant to have been so arranged (as is always possible) that

$$r_{11} \geqslant r_{12} \geqslant \cdots \geqslant r_{1n}.$$

The converse of the two-factor theorem now states that it is possible so to determine C, g and s that, for the given N individuals, and for all mutually different subscripts k, l,

$$r_{s_k s_l} = 0 \quad \text{...(i)},$$

$$r_{s_k g} = 0 \quad \text{...(ii)},$$

and this is what we are now required to show.

Proof. There exists by hypothesis an n-way correlation solid, that is, a distribution in n dimensional space, determined by the n aptitudes $a_{\alpha,\beta}$, $\alpha = 1, 2, \ldots, n$, of the N individuals, and the total correlations in this solid are as indicated in Δ. It is necessary to assume that Δ is not zero, but this is a trivial assumption, for the vanishing of this determinant would mean that the n total regression "planes" of this solid would have a "line" in common, instead of the general mean point only, and in other geometrical ways this distribution would be so peculiar as to be unrealisable in practice. We shall now show that there exists also a distribution of these N individuals in $(n+1)$-way space, determined by the given N values of each of the n variables $a_1, \ldots, a_n$, and by the N values to be assigned to a single new variable g; thus the mean g is zero, the total correlations are as indicated in the determinant R of (iv), and also, for R, condition (c) holds, even when the leading element $r_{gg}=1$ is included, and thus, finally, for all mutually different values of k and l,

$$r_{kl} = r_{gl} \cdot r_{gk} \qquad \text{(iii)},$$

$$R = \begin{vmatrix} r_{gg} & r_{g1} & r_{g2} & \cdots & r_{gn} \\ r_{1g} & r_{11} & r_{12} & \cdots & r_{1n} \\ r_{2g} & r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{ng} & r_{n1} & r_{n2} & \cdots & r_{nn} \end{vmatrix} \qquad \text{(iv)}.$$

First, such a set of positive numbers does exist, and none are greater than unity; for, put $r_{gg}=1$, and let r_{g1}, r_{g2} and r_{g3} be determined by the equations

$$\left.\begin{aligned} r_{12} &= r_{g1} r_{g2} \\ r_{13} &= r_{g1} r_{g3} \\ r_{23} &= r_{g2} r_{g3} \end{aligned}\right\} \qquad \text{(v)}.$$

The solution of (v) is

$$r_{g1} = \sqrt{\frac{r_{12} r_{13}}{r_{23}}}, \quad r_{g2} = \sqrt{\frac{r_{12} r_{23}}{r_{13}}}, \quad r_{g3} = \sqrt{\frac{r_{13} r_{23}}{r_{12}}},$$

and hence by (b) the three numbers r_{g1}, r_{g2}, r_{g3} are positive and not greater than 1. Then, if $l>3$, let

$$r_{gl} = r_{1l} \cdot \frac{1}{r_{1g}} = \sqrt{\frac{r_{1l}}{r_{13}} \cdot \frac{r_{1l}}{r_{12}}} \cdot r_{23}.$$

Since $r_{1l} \leqslant r_{12}$ and $r_{1l} \leqslant r_{13}$ and $r_{23} \leqslant 1$, it follows that

$$r_{gl} \leqslant 1 \qquad \text{(vi)}.$$

It also follows from these equations that

$$r_{g1} = \sqrt{\frac{r_{k1} r_{l1}}{r_{kl}}}, \text{ etc.},$$

for all mutually different values of k and l, and so, adopting here the notation of Piaggio and others, we have

$$r_{g1} = \frac{1}{\mu_1}, \ \ldots, \ r_{gn} = \frac{1}{\mu_n} \qquad \text{(vii)}.$$

All these interrelationships may be represented compactly by the following matrix, M, in which parallel arrays are proportional, now without the exception of the principal diagonal:

$$M: \left\| \begin{array}{cccccc} 1 & r_{g1} & r_{g2} & r_{g3} & \dots & r_{gn} \\ r_{1g} & r^2_{g1} & r_{12} & r_{13} & \dots & r_{1n} \\ r_{2g} & r_{21} & r^2_{g2} & r_{23} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{ng} & r_{n1} & r_{n2} & r_{n3} & \dots & r^2_{gn} \end{array} \right\| \qquad \text{(viii)}.$$

So far, we have proved that there actually does exist a set of numbers as indicated in M, and that each of them is positive and at most equal to unity, but we have not shown that they can all be correlation coefficients. That depends on whether there exists an $(n+1)$-way frequency distribution whose total correlations are these numbers, and this is the important thing it now remains for us to establish. A simple but rather helpful picture of the method is afforded by a case of three variables. Consider a two-way frequency distribution between x and y. This may be represented by a set of points on a horizontal (xy) plane. Of course r_{xy} is known. Now suppose we have the problem of constructing a three-way distribution in xyz space, having the same r_{xy} and also some pre-assigned r_{yz} and r_{zx}. It is clear that we may move the points vertically at pleasure without disturbing x or y or r_{xy}. If we do this in such a way as to produce the proper r_{yz} when the points are projected on the yz plane, it may be that we will not thus have produced the proper r_{zx} when they are projected on the zx plane. The question whether we can do both at once depends on how many points there are and how many distinct conditions must be fulfilled; and the most direct way of solving the problem would seem to be, therefore, to write out the conditions and count them. This is what we now propose to do, but our points exist initially in space of n dimensions instead of two. The $(n+1)$th direction is the direction of g. The number of points is N. Each of these is to be moved in the g direction to its proper position, and the co-ordinate of that new position is also to be called g. The co-ordinates of the N points, as located ultimately, are subject to n conditions only, viz.

$$\frac{1}{N}\sum_{\beta} g a_i = \sigma_g r_{gi}, \quad (i = 1, 2, \dots, n) \qquad \text{(ix)},$$

where

$$\sigma_g^2 = \frac{1}{N}\sum_{\beta} g^2,$$

and from them the function g is to be determined. Since there are n conditions, the most natural function of the a's to assume as the form of g is the linear combination

$$y = A_1 a_1 + \dots + A_n a_n \qquad \text{(x)},$$

but this is by no means necessary. Any other function involving n arbitrary constants might do as well, or better, and we shall see later that we may start with vastly more general functions. It happens too that this linear function is not quite

sufficient by itself, because when inserted in (ix) it would lead to n equations in the A's which would not have a solution. Due to the presence of σ_g, these equations are not linear in the A's. In order to meet this new difficulty, it is necessary to examine it more closely. Solve Equations (ix) inserting $g=y$,

$$\frac{1}{N}\sum_{\beta}(A_1a_1+\ldots+A_na_n)\,a_i=\sigma_g r_{gi},\quad (i=1,2,\ldots,n),$$

first as if σ_g were a known number, noting that they become

$$\left.\begin{array}{l} A_1r_{11}+A_2r_{12}+\ldots+A_nr_{1n}=r_{g1}\sigma_g \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ A_1r_{n1}+A_2r_{n2}+\ldots+A_nr_{nn}=r_{gn}\sigma_g \end{array}\right\}\quad\ldots\ldots\ldots\ldots\ldots\ldots(\text{xi}).$$

Due to the proportional relationships that exist among the r's,

$$\Delta=\frac{(\mu_1^2-1)\ldots(\mu_n^2-1)}{K^2\mu_1\ldots\mu_n},$$

and the solution is $$A_i=K^2\sigma_g\frac{\mu_i}{\mu_i^2-1},\quad (i=1,2,\ldots,n)\ \ldots\ldots\ldots(\text{xii}),$$

where $$K^2=\frac{1}{1+\dfrac{1}{\mu_1^2-1}+\ldots+\dfrac{1}{\mu_n^2-1}}.$$

This is not a true solution, because σ_g is a function of the A's, but (xii) is a more illuminating form in which (xi) may be written. Together with (xii) we must now satisfy also the equation $\sigma_g=\sigma_y$, where, summing for all individuals:

$$\sigma_y^2=\frac{1}{N}\Sigma\,(A_1a_1+\ldots+A_na_n)^2.\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{xii}\,a)$$

It will be shown that this cannot be done, except in the trivial case where all the A's vanish. By (xii$a$),

$$\sigma_y^2=(A_1^2+\ldots+A_n^2)+2\,(A_1A_2r_{12}+\ldots+A_{n-1}A_nr_{n-1,n}),$$

and by (xii),

$$A_iA_jr_{ij}=K^4\sigma_g^2\frac{\mu_i\mu_jr_{ij}}{(\mu_i^2-1)(\mu_j^2-1)}=\frac{K^4\sigma_g^2}{(\mu_i^2-1)(\mu_j^2-1)},$$

and so $$\sigma_y^2=K^4\sigma_g^2\left\{\left(\frac{\mu_1}{\mu_1^2-1}\right)^2+\ldots+\left(\frac{\mu_n}{\mu_n^2-1}\right)^2\right.$$
$$\left.+2\left(\frac{1}{(\mu_1^2-1)(\mu_2^2-1)}+\ldots+\frac{1}{(\mu_{n-1}^2-1)(\mu_n^2-1)}\right)\right\}$$
$$=\sigma_g^2(1-K^2)\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{xiii}).$$

Therefore, we cannot satisfy both (xii), which is the same as (xi), and the condition that $\sigma_y^2=\sigma_g^2$, i.e. we cannot let $g=y$, unless $K=0$. The extreme case where K is zero, or near it, is discussed in Section 5. It is the condition for an "almost" unique g.

Having moved our points up to the position occupied by the plane

$$y=A_1a_1+\ldots+A_na_n\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{xiv}),$$

and having found that they satisfied all the requirements, except that the standard deviation was too small, the obvious thing to do next would be to move them away from this plane so as to secure the correct standard deviation but not disturb what has already been accomplished. So we make this plane the regression plane of the points in their final position, denoting by t their distances (parallel to g) from it. This gives us a good deal of latitude in our choice of t's, but we are restricted somewhat. Let $g = t + y$, where, for each i,

$$\sum_\beta t = \sum_\beta t a_i = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xv)}.$$

Then the condition (ix) will yield (xi) as before, and

$$\sigma_g^2 = \sigma_t^2 + \sigma_y^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xvi)},$$

and so we must choose t so that both (xv) and (xvi) will be satisfied; then (xii) will yield the solution. This may not be the only way in which one may obtain a g which will have the necessary requirements, but it is certainly one way. By (xiii), condition (xvi) is the same as

$$\sigma_g^2 = \sigma_t^2 + \sigma_g^2 (1 - K^2),$$

and hence

$$\sigma_t = K\sigma_g \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xvii)}.$$

This choice of σ_t is arbitrary in so far as σ_g is arbitrary, but in order that g may be compared with the a's it is desirable to have $\sigma_g = 1$, and so $\sigma_t = K$, and

$$\sigma_y = \sqrt{1 - K^2}.$$

Thus there is a definite restriction on the freedom of t, and, as stated earlier, it has to do with the almost uniqueness of g. Postponing that subject, we now complete the proof. It proceeds from here as indicated by Spearman and others, and so may be indicated very briefly. It remains only to show that (i) and (ii) are satisfied It follows from (iii) that the net correlation, $r_{kl\cdot g} = 0$, if k and l are different. The total regressions of a_k and a_l on g are: $a_k = r_{gk} g$ and $a_l = r_{gl} g$, and it is known from general correlation theory that this net correlation is the simple correlation between $(a_k - r_{gk} g)$ and $(a_l - r_{gl} g)$. Set these equal to s_k and s_l respectively, and let $C_a = r_{ga}$, and it follows that we have so chosen the variables of (a) § 3 that (i) is satisfied:

$$r_{s_k s_l} = 0,\ k \neq l.$$

To obtain (ii), compute the correlation between s_k and g. Since

$$\sum_\beta (a_k - g r_{gk}) = 0 \ \text{ and } \ \sigma_{s_k}^2 = \frac{1}{N} \sum_\beta (a_k - g r_{gk})^2 = 1 - r_{gk}^2 \quad \ldots\ldots\text{(xviii)},$$

$$r_{s_k g} = \frac{1}{N\sqrt{1 - r_{gk}^2}} \sum_\beta (a_k - g r_{gk})\, g = \frac{1}{\sqrt{1 - r_{gk}^2}} (r_{gk} - r_{gk}) = 0,$$

as desired.

4. Necessity. We shall now show the necessity of the condition in (b) $r_{ij\cdot k} \geqslant 0$, namely, that it follows from the two-factor theorem and from this theorem only that $r_{ij} \geqslant r_{ik} r_{jk}$. It is quite well known that it follows from the two-factor theorem in the direct (not the converse) form that

$$r_{gk}^2 = \frac{r_{ik} r_{jk}}{r_{ij}},$$

where r_{gk} is a correlation coefficient which is positive and at most equal to unity; so $r_{ij} \geqslant r_{ik} r_{jk}$.

5. Interpretation. As stated in Section 3, the values of g may be assigned to the N individuals in many ways, at least provided N is greater than $n+1$. This means that we may choose, for example, for the general factor applicable to the first individual a value twice as great or half as great as the one assigned to the second individual, at pleasure, and still meet all the conditions of this theorem. This sort of a general factor is not, I should think, what psychologists desire in order to establish the two-factor theory of mental abilities. If the number of ways in which one may assign values of the general factor to the N individuals is very large instead of unique, it would seem to me doubtful whether from a psychological point of view it would be meaningful to assert the existence of such a "factor" at all, but that is obviously a question for psychologists. That it is true, however, that the general factor is strikingly indeterminate is well established in the numerical example of Section 6. Consider the 11th and the 12th individuals ($\beta = 11, 12$) of that example. For the 11th, the general factor equals 0·85 for one choice of t's, $-1·77$ for another; for the 12th, these figures are reversed. The scale (σ) of these measurements is unity and the origins are at their means. The disparity is perhaps the more striking because both these individuals have been assigned the same scores on all four of the tests. It would not be relevant to assert that I had made a peculiar choice of t's in order to secure these differences, that a random choice would probably not have produced them, for the choice of t's is not restricted to be random; it is arbitrary. The so-called arbitrary constant of integration affords a perfect analogy. Given an hierarchy, one can assert the existence of a general factor in exactly the same sense as, when given an analytic function, one can assert the existence of its anti-derivative or "indefinite" integral; there will exist a family of general factor functions, every member of which would lead to the given hierarchy.

Admitting this, Piaggio and Irwin have asserted that nevertheless the general factor g is what one might call "almost" unique, in the sense that the possible fluctuations of g from y are small on the average, at least in certain circumstances. I shall discuss this fluctuation in a moment, but let us consider first a possible demurrer. If their contention is justified, it means only that, for any choice of t's, the average deviation of g from y is small; each member of the infinite family of surfaces of which y is the common regression plane lies on the average close to that plane. Not every point of it needs to be close to that plane; whatever individual be selected, it is possible to find at least one member of that family for which his g will differ greatly (if N be much greater than n) from y. For any given hierarchy, then, we may write down a whole family of two-factor patterns, some of which will differ widely from each other in the case of any previously selected individual. However, if the contention of these authors be justified, it does follow that for each member of that family the majority of the individuals must have g's close to y. We do know that our hierarchy must have associated with it a set of g's which deter-

mine one or another member of this family, and this set as a whole lies close to y. Therefore, the group behaviour of the N individuals might be expected to be as if g were unique. So an investigation of the fluctuation of g is warranted.

We have seen that, if $\sigma_g = \sigma_a = 1$, then $\sigma_t = K$, $\sigma_y = \sqrt{1-K^2}$, $g = t + y$. It is a question whether the approximation, $g = y$, is a close one, and this depends on whether $(\sigma_t = K)$ is small, relative to $(\sigma_g = 1)$, say as small as 0·1. In terms of the r's,

$$\frac{1}{K^2} = 1 + \frac{r_{1g}^2}{1 - r_{1g}^2} + \ldots + \frac{r_{ng}^2}{1 - r_{ng}^2}.$$

There are two ways in which $1/K$ can be as large as 10. One is when the early r's are very close to 1. From the equation,

$$a_{1\beta} = r_{g1} g_\beta + s_{1\beta},$$

we get $$\sum_\beta a_1^2 = r_{g1} \sum_\beta g a_1 + \sum_\beta s_1 a_1.$$

Hence, using (xviii), $\quad 1 = r_{g1}^2 + r_{s_1 a_1} \sigma_{s_1}, \quad r_{s_1 a_1} = \sqrt{1 - r_{g1}^2}.$

So, if r_{g1} is close to 1, $r_{s_1 a_1}$ is close to zero. Moreover, for values of k different from 1 it follows from the earlier results (§ 3, i, and ii), that $r_{s_k a_1} = 0$ exactly. Therefore, a_1 has by hypothesis all the properties of the general factor sought, approximately. In other words, this is a trivial case where the g of the two-factor theorem has already been found, approximately, among the aptitudes measured.

The other case is where n is so large that, even if each term of $1/K^2$ is small, yet their sum is large. For example, if $r_{kl} = 0·5$, for all $k \neq l$, $r_{gk}^2 = 0·5$, and

$$\frac{1}{K^2} = 1 + 1 + \ldots + 1,$$

so that $1/K = 10$ if $n = 99$. To accomplish the same result when $r_{kl} = 0·25$ would require the measurement of 297 aptitudes. Practically, it would appear difficult to secure cases of hierarchies among aptitudes as numerous as this with intercorrelations as large, but if they should occur, it would then be true that the *group* general factor would be "almost" unique. It is to be remembered that for the success of our theory the number of individuals N must exceed the number of tests n*.

* The relationship between Piaggio's notation and mine is as follows. His $a = m_a g + n_a s_a$, where $\sigma_a = 1$, $\sigma_g = 1$, $\sigma_s = 1$. My $a_a = C_a g + s_a$, where $\sigma_a = 1$, $\sigma_g = 1$, $\sigma_{s_a} = \sqrt{1 - r_{ga}^2}$. His $g = k^2 t + ki$, where

$$\sigma_t = \frac{\sqrt{1-k^2}}{k^2}, \quad \sigma_i = 1.$$

My $g = y + t$, where $\sigma_y = \sqrt{1 - K^2}$, $\sigma_t = K$. His μ = my μ. His k = my K. Thus his indeterminate part of g is ki and mine is t. Now he says that by increasing his N (my n), the coefficient of the uncertainty term, k, "can be made as small as we please," and concludes from that that g may become almost unique. This does not seem conclusive, for if his k is less than 1, his k^2 is still smaller and that is the coefficient of the determined term (his t, my y). Irwin rightly judges that it is a matter of standard deviations as well as of coefficients, but he says that k^{-2} can be made small at pleasure. This looks like a misprint for k^2, for, as noted before, it is the smallness of k^2 which is needed to render g almost unique.

According to Wilson, any set of n aptitudes which do not lead to a hierarchy may be replaced by n artificial aptitudes, which are certain combinations of the given aptitudes, which do lead to a hierarchy. Since n may be as large as desired, it would follow that there always exists a set of n artificial aptitudes for which there is a group general factor which is almost unique.

This argument for uniqueness has proceeded from the assumption that the form of the function g was $g = y + t$, where y was linear and t had special characteristics, but it was remarked that it was not necessary, initially, that y should be linear. Indeed, returning to (x), we might have written $g = \phi + t$, where

$$\phi = A_1 f_1 + \dots + A_n f_n,$$

and f_i was any function of $a_{i,\beta}$ such that

$$\sum_\beta f_i = 0 \quad \text{and} \quad \sum_\beta f_i^2 = N, \qquad (i = 1, \dots, n).$$

In that case, values of A_i could have been found, usually, subject to the same conditions (ix) as before, and from that point on the original proof would have held good. But, nevertheless, in that case, the set of values found for g would still have constituted some sort of a frequency distribution in $(n+1)$ space and this distribution would have had a regression plane. Also, it is known from general correlation theory that the differences between the ordinates to such a plane and the values of g thus found would have had the properties previously ascribed to t:

$$\Sigma t = 0, \quad \Sigma a_i t = 0.$$

Hence this new g could be represented as the sum of a linear function and such a t; and so it has now been shown that our original solution was in fact the most general one possible.

6. Example. For the individuals $\beta = 1$ to 40, let the a's be the observations, having the values indicated in the table. It follows that the determinant R (see iv) is as below, indicating a hierarchy of r_{ij}'s and the r_{gi}'s associated with it. The y's are determined from the equations

$$y = \frac{6}{7\sqrt{5}}(a_1 + a_3) + \frac{2}{5\sqrt{7}}(a_2 + a_4),$$

$$A_1 = \frac{6}{7\sqrt{5}}, \quad A_2 = \frac{2}{5\sqrt{7}}, \quad A_3 = A_1, \quad A_4 = A_2. \quad K^2 = \frac{3}{35}, \quad N = 40, \quad n = 8.$$

β	a_1	a_2	a_3	a_4	y	
1—5	$2c$	$2c$	c	c	$12c$	$-1{\cdot}37$
6—10	0	$2c$	c	c	$6c$	$\cdot 69$
11—15	$-c$	$-c$	0	0	$-4c$	$-\cdot 46$
16—20	$-c$	$-c$	$-2c$	0	$-10c$	$-1{\cdot}14$
21—25	$-c$	$-c$	$-c$	$-3c$	$-10c$	$-1{\cdot}14$
26—30	c	$-c$	c	c	$6c$	$\cdot 69$
31—35	c	c	c	c	$8c$	$\cdot 91$
36—40	$-c$	$-c$	$-c$	$-c$	$-8c$	$-\cdot 91$
Σ	0	0	0	0	0	$\cdot 01$
σ	1	1	1	1		$\cdot 91$
c	$2/\sqrt{5}$	$2/\sqrt{7}$	$2/\sqrt{5}$	$2/\sqrt{7}$	$4/35$	

$$R = \begin{vmatrix} 1 & \frac{2}{\sqrt{5}} & \frac{2}{\sqrt{7}} & \frac{2}{\sqrt{5}} & \frac{2}{\sqrt{7}} \\ \frac{2}{\sqrt{5}} & 1 & \frac{4}{\sqrt{35}} & \frac{4}{5} & \frac{4}{\sqrt{35}} \\ \frac{2}{\sqrt{7}} & \frac{4}{\sqrt{35}} & 1 & \frac{4}{\sqrt{35}} & \frac{4}{7} \\ \frac{2}{\sqrt{5}} & \frac{4}{5} & \frac{4}{\sqrt{35}} & 1 & \frac{4}{\sqrt{35}} \\ \frac{2}{\sqrt{7}} & \frac{4}{\sqrt{35}} & \frac{4}{7} & \frac{4}{\sqrt{35}} & 1 \end{vmatrix}.$$

The t's must now be chosen so as to satisfy the relations (xv) and the condition that $\sigma_t^2 = 3/35$. Two simple choices are:

Case I: $t = \sqrt{12/7} = 1{\cdot}31$ for $\beta = 11$, and $t = -1{\cdot}31$ for $\beta = 12$, $t = 0$ for all other β's.

Case II: $t = -1{\cdot}31$ for $\beta = 11$, and $t = 1{\cdot}31$ for $\beta = 12$, $t = 0$ otherwise. Thence $g = y + t$, and then the s's are determined so as to satisfy the following equations for each β:

$$a_1 = 2g/\sqrt{5} + s_1, \quad a_2 = 2g/\sqrt{7} + s_2,$$
$$a_3 = 2g/\sqrt{5} + s_3, \quad a_4 = 2g/\sqrt{7} + s_4.$$

Some of the values of g and s_1 are approximately as in the next table.

β	Case I		Case II	
	g	s_1	g	s_1
10	·69	− ·62	·69	− ·62
11	·85	−1·65	−1·77	·69
12	−1·77	·69	·85	−1·65
13	− ·46	− ·48	− ·46	− ·48
σ	1	·45	1	·45

THE DISTRIBUTION OF THE INDEX IN A NORMAL BIVARIATE POPULATION.

By E. C. FIELLER, B.A.

The Probability Integral of the Index Distribution.

Consider the distribution of the ratio

$$v = \frac{y}{x}$$

in *any* bivariate population $\qquad z = f(x, y) \quad \ldots\ldots\ldots\ldots\ldots\ldots(1),$

where $$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\,dx\,dy = 1.$$

Points (x, y) corresponding to a given value of v lie on the line

$$y = vx \quad \ldots\ldots\ldots\ldots\ldots\ldots(2);$$

hence the chance that a random member of the population (1) will have an index lying in the range $v_1 \leqslant v \leqslant v_2$ is equal to the volume of the portion of the frequency surface (1) that lies above the area swept out in the xy-plane by the line (2) as it revolves in the positive direction from the position

$$y = v_1 x \quad \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

to the position $$y = v_2 x \quad \ldots\ldots\ldots\ldots\ldots\ldots(4).$$

If we take $v_1 = -\infty$, this volume is the chance that the index will not exceed v_2; the line (3) is then the y-axis, and the volume is

$$V_2 = \int_0^{\infty}\int_{-\infty}^{v_2 x} + \int_{-\infty}^{0}\int_{v_2 x}^{\infty} f(x, y)\,dx\,dy \quad \ldots\ldots\ldots\ldots\ldots\ldots(5).$$

When the joint distribution of x and y is the normal one,

$$z = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\,e^{-\frac{1}{2}\frac{1}{1-r^2}\left\{\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2r\frac{x-\bar{x}}{\sigma_x}\frac{y-\bar{y}}{\sigma_y} + \left(\frac{y-\bar{y}}{\sigma_y}\right)^2\right\}} \quad \ldots\ldots(6),$$

(5) gives, for the chance of an index not greater than v,

$$V = \int_{a+b}\int \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\,e^{-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{x^2}{\sigma_x^2} - 2r\frac{x}{\sigma_x}\frac{y}{\sigma_y} + \frac{y^2}{\sigma_y^2}\right)}\,dx\,dy \quad \ldots\ldots(7),$$

where a and b are the two portions of the xy-plane indicated in Fig. 1.

The boundaries of a and b are the lines

$$x + \bar{x} = 0,$$

$$y + \bar{y} = v(x + \bar{x}),$$

and if we put $x = \xi$, $y - vx = \eta$, the portions α and β of the $\xi\eta$-plane that correspond to a and b are bounded (see Fig. 2) by the lines

$$\xi + \bar{x} = 0,$$
$$\eta + \bar{y} - v\bar{x} = 0.$$

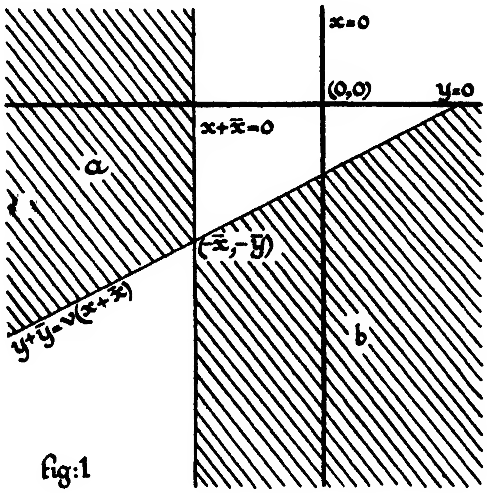


fig:1

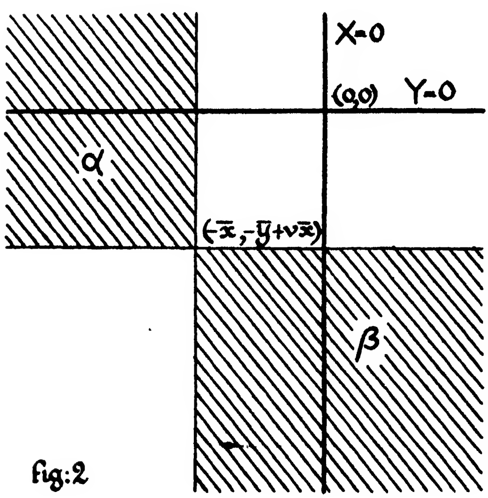


fig:2

Thus we have, performing the change of variables,

$$V = \int_{\alpha+\beta}\int \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} e^{-\frac{1}{2}\chi^2}\, d\xi\, d\eta \quad \text{.....................(8)},$$

where
$$\chi^2 = \frac{1}{1-r^2}\left\{\frac{\xi^2}{\sigma_x^2} - 2r\frac{\xi}{\sigma_x}\frac{\eta+v\xi}{\sigma_y} + \left(\frac{\eta+v\xi}{\sigma_y}\right)^2\right\}$$
$$\equiv \frac{1}{1-\rho^2}\left\{\frac{\xi^2}{\sigma_\xi^2} - 2\rho\frac{\xi}{\sigma_\xi}\frac{\eta}{\sigma_\eta} + \frac{\eta^2}{\sigma_\eta^2}\right\}, \text{ say} \quad \text{.......................(9)}.$$

From the last identity we have

$$\sigma_\xi^2(1-\rho^2)=\sigma_x^2\sigma_y^2(1-r^2)/(\sigma_y^2-2rv\sigma_x\sigma_y+v^2\sigma_x^2),$$

$$\sigma_\eta^2(1-\rho^2)=\sigma_y^2(1-r^2),$$

$$\frac{1}{\sigma_\xi\sigma_\eta}\frac{\rho}{1-\rho^2}=(r\sigma_y-v\sigma_x)/\sigma_x\sigma_y^2(1-r^2);$$

squaring the last of these equations, and multiplying by the other two, we get

$$\rho^2=(r\sigma_y-v\sigma_x)^2/(\sigma_y^2-2rv\sigma_x\sigma_y+v^2\sigma_x^2)\quad\quad(10),$$

whence

$$1-\rho^2=(1-r^2)\,\sigma_y^2/(\sigma_y^2-2rv\sigma_x\sigma_y+v^2\sigma_x^2)\quad\quad(11),$$

$$\sigma_\xi=\sigma_x\quad\quad(12),$$

$$\sigma_\eta=(\sigma_y^2-2rv\sigma_x\sigma_y+v^2\sigma_x^2)^{\frac{1}{2}}\quad\quad(13),$$

and

$$\sigma_\xi\sigma_\eta\sqrt{1-\rho^2}=\sigma_x\sigma_y\sqrt{1-r^2}\quad\quad(14).$$

Write

$$X=\frac{\xi}{\sigma_\xi}=\frac{x}{\sigma_x},$$

$$Y=\frac{\eta}{\sigma_\eta}=\frac{y-vx}{(\sigma_y^2-2rv\sigma_x\sigma_y+v^2\sigma_x^2)^{\frac{1}{2}}};$$

the quadrants A and B of the XY-plane that correspond to the portions a and b of the xy-plane have as common corner the point $(-h, -k)$, where

$$h=\frac{\bar{x}}{\sigma_x}\quad\quad(15),$$

$$k=\frac{\bar{y}-v\bar{x}}{(\sigma_y^2-2rv\sigma_x\sigma_y+v^2\sigma_x^2)^{\frac{1}{2}}}\quad\quad(16).$$

From (8),

$$V=\int_{A+B}\int\frac{\sigma_\xi\sigma_\eta}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}e^{-\frac{1}{2}\frac{1}{1-\rho^2}(X^2-2\rho XY+Y^2)}\,dX\,dY,$$

so that the chance of obtaining an index not less than v is

$$C=1-V=\int_h^\infty\int_k^\infty+\int_{-h}^\infty\int_{-k}^\infty\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{1-\rho^2}(X^2-2\rho XY+Y^2)}\,dX\,dY\quad(17).$$

Here h and k are given by (15) and (16); equation (10) provides two values of ρ; we decide which is appropriate by noting that as $v\to\infty$, the point

$$(h, k)\to\left(\frac{\bar{x}}{\sigma_x}, -\frac{\bar{x}}{\sigma_x}\right),$$

so that to make $1-V\to 0$ we must take

$$\rho=\frac{r\sigma_y-v\sigma_x}{(\sigma_y^2-2rv\sigma_x\sigma_y+v^2\sigma_x^2)^{\frac{1}{2}}}\quad\quad(18).$$

Tables are already in existence by means of which the numerical value of C may be found*. These tables show over a range of values of ρ extending from -1 to $+1$, the value of

$$\int_h^\infty\int_k^\infty Z(\rho)\,dX\,dY=\int_h^\infty\int_k^\infty\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{1-\rho^2}(X^2-2\rho XY+Y^2)}\,dX\,dY$$

* *Tables for Statisticians and Biometricians*, Part II, Tables VIII and IX.

for positive values of h and k; for evaluating the integral when h and k are not both positive, we use the relations

$$\int_{-h}^{\infty}\int_{k}^{\infty} Z(\rho)\,dX\,dY = \int_{k}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Y^2}\,dY - \int_{h}^{\infty}\int_{k}^{\infty} Z(-\rho)\,dX\,dY,$$

$$\int_{h}^{\infty}\int_{-k}^{\infty} Z(\rho)\,dX\,dY = \int_{h}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}X^2}\,dX - \int_{h}^{\infty}\int_{k}^{\infty} Z(-\rho)\,dX\,dY,$$

$$\int_{-h}^{\infty}\int_{-k}^{\infty} Z(\rho)\,dX\,dY = 1 - \int_{h}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}X^2}\,dX - \int_{k}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Y^2}\,dY + \int_{h}^{\infty}\int_{k}^{\infty} Z(\rho)\,dX\,dY.$$

Thus if the h and k provided by (15) and (16) be both positive or both negative, we have

$$C = 1 - \overline{\int_{|h|}^{\infty} + \int_{|k|}^{\infty}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx + 2\int_{|h|}^{\infty}\int_{|k|}^{\infty} Z(\rho)\,dX\,dY \quad\text{......(17 }a);$$

while if h and k be of opposite sign,

$$C = \int_{|h|}^{\infty} + \int_{|k|}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx - 2\int_{|h|}^{\infty}\int_{|k|}^{\infty} Z(-\rho)\,dX\,dY \quad\text{.........(17 }b)$$

Frequency Distribution of the Index.

By differentiating V with respect to v, we have the frequency distribution of v. Since h does not vary with v,

$$\frac{dV}{dv} = \frac{\partial V}{\partial \rho}\frac{\partial \rho}{\partial v} + \frac{\partial V}{\partial k}\frac{\partial k}{\partial v} \quad\text{..............................(19).}$$

By a well-known property of the normal surface,

$$\frac{\partial}{\partial\rho}\left\{\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{1-\rho^2}(X^2-2\rho XY+Y^2)}\right\} = \frac{\partial^2}{\partial X\,\partial Y}\left\{\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{1-\rho^2}(X^2-2\rho XY+Y^2)}\right\},$$

so that (17) gives

$$-\frac{\partial V}{\partial \rho} = 2\,\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{1-\rho^2}(h^2-2\rho hk+k^2)}$$

$$= \frac{1}{\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{\bar{x}^2}{\sigma_x^2}-2r\frac{\bar{x}}{\sigma_x}\frac{\bar{y}}{\sigma_y}+\frac{\bar{y}^2}{\sigma_y^2}\right)} \quad\text{..................(20)}$$

by (15), (16), and (9).

From (11),
$$-\rho\frac{\partial\rho}{\partial v} = \frac{\sigma_x\sigma_y^2(1-r^2)(r\sigma_y - v\sigma_x)}{(\sigma_y^2 - 2rv\sigma_x\sigma_y + \sigma_x^2)^2},$$

which, with (11), (18), and (20), gives

$$\frac{\partial V}{\partial\rho}\frac{\partial\rho}{\partial v} = \frac{\sigma_x\sigma_y\sqrt{1-r^2}}{\pi(\sigma_y^2 - 2rv\sigma_x\sigma_y + \sigma_x^2)}e^{-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{\bar{x}^2}{\sigma_x^2}-2r\frac{\bar{x}}{\sigma_x}\frac{\bar{y}}{\sigma_y}+\frac{\bar{y}^2}{\sigma_y^2}\right)} \quad\text{......(21).}$$

From (16) we find

$$(\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2)^{\frac{3}{2}} \frac{\partial k}{\partial v} = \sigma_y (r\bar{y}\sigma_x - \bar{x}\sigma_y) + v\sigma_x (r\bar{x}\sigma_y - \bar{y}\sigma_x) \quad \text{......(22)},$$

and from (17)

$$\frac{\partial V}{\partial k} = \frac{1}{2\pi\sqrt{1-\rho^2}} \left\{ \int_h^\infty e^{-\frac{1}{2}\frac{1}{1-\rho^2}(x^2 - 2\rho xk + k^2)} dx - \int_{-h}^\infty e^{-\frac{1}{2}\frac{1}{1-\rho^2}(x^2 + 2\rho xk + k^2)} dx \right\}$$

$$= \frac{1}{2\pi} e^{-\frac{1}{2}k^2} \left\{ \int_{\frac{h-\rho k}{\sqrt{1-\rho^2}}}^\infty e^{-\frac{1}{2}u^2} du - \int_{-\frac{h-\rho k}{\sqrt{1-\rho^2}}}^\infty e^{-\frac{1}{2}u^2} du \right\}$$

$$= \frac{1}{\pi} e^{-\frac{1}{2}k^2} \int_0^{\frac{\rho k - h}{\sqrt{1-\rho^2}}} e^{-\frac{1}{2}u^2} du \quad \text{..(23)}.$$

Inserting the values of h, k, and ρ in (23), we have, using (20), (21), and (22), the frequency distribution of v:

$$\psi(v) = \frac{1}{\pi} \frac{\sigma_x\sigma_y\sqrt{1-r^2}}{\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2} e^{-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{\bar{x}^2}{\sigma_x^2} - 2r\frac{\bar{x}}{\sigma_x}\frac{\bar{y}}{\sigma_y} + \frac{\bar{y}^2}{\sigma_y^2}\right)}$$

$$+ e^{-\frac{1}{2}\frac{(\bar{y} - v\bar{x})^2}{\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2}} \frac{\sigma_y (r\bar{y}\sigma_x - \bar{x}\sigma_y) + v\sigma_x (r\bar{x}\sigma_y - \bar{y}\sigma_x)}{\pi(\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2)^{\frac{3}{2}}}$$

$$\times \int_0^{\frac{\sigma_y (r\bar{y}\sigma_x - \bar{x}\sigma_y) + v\sigma_x (r\bar{x}\sigma_y - \bar{y}\sigma_x)}{\sigma_x\sigma_y\{(1-r^2)(\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2)\}^{\frac{1}{2}}}} e^{-\frac{1}{2}u^2} du \quad \text{.........(24)}.$$

We can obtain this distribution in a somewhat more direct manner. If we write

$$x = x, \quad v = \frac{y}{x}, \quad \left| \frac{\partial (xv)}{\partial (xy)} \right| = |x| \quad \text{.....................(25)},$$

we have, from (6), the equation to the joint distribution of x and v:

$$\phi(x, v) = \frac{|x|}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} e^{-\frac{1}{2}\frac{1}{1-r^2}\left\{\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2r\frac{x-\bar{x}}{\sigma_x}\frac{vx-\bar{y}}{\sigma_y} + \left(\frac{xv-\bar{y}}{\sigma_y}\right)^2\right\}} \quad \text{...(26)}.$$

On integrating this equation with respect to x, we arrive at (24). As we should expect, (24) is not altered if we increase $\bar{x}$, $\bar{y}$, σ_x, and σ_y in the same ratio *.

Distribution of the Index in a Curtailed Normal Population.

The two terms of which the second member of (24) is the sum are essentially positive; accordingly, the moments of the index-distribution $\psi(v)$ will be infinite with the contributions of the first of these terms. It is obvious that these infinities would not arise if we restricted (x, y) to some limited region in the positive quadrant, since then the index-distribution would have a limited range.

[* An erroneous solution of this problem having been sent to me by Mr G. A. Baker, the above solution (24) by Mr Fieller was forwarded to him with permission to use it. The result in Equation (24) was published by Mr Baker in *The Annals of Mathematical Statistics*, Vol. III. p. 5, February, 1932, the mention of Mr Fieller being unfortunately overlooked. Ed.]

Suppose, for example, that we take the joint distribution of x and y to be

$$z = z_0 e^{-\frac{1}{2}\frac{1}{1-r^2}\left\{\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2r\frac{x-\bar{x}}{\sigma_x}\frac{y-\bar{y}}{\sigma_y} + \left(\frac{y-\bar{y}}{\sigma_y}\right)^2\right\}} \quad \text{..........(27)},$$

provided that (x, y) lies inside the ellipse

$$\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2r\frac{x-\bar{x}}{\sigma_x}\frac{y-\bar{y}}{\sigma_y} + \left(\frac{y-\bar{y}}{\sigma_y}\right)^2 = \lambda^2 \quad \text{..........(28)},$$

which is a probability contour of the normal surface (6), and zero if (x, y) lie outside this ellipse.

Then by applying the transformation (25) and integrating out with respect to x, we have, for the ordinate of the curve of distribution of v,

$$I(\lambda) = z_0 \int_{x_1(v)}^{x_2(v)} |x| \, e^{-\frac{1}{2}\frac{1}{1-r^2}\left\{\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2r\frac{x-\bar{x}}{\sigma_x}\frac{vx-\bar{y}}{\sigma_y} + \left(\frac{vx-\bar{y}}{\sigma_y}\right)^2\right\}} dx \quad \text{......(29)},$$

where $x_1(v)$ and $x_2(v)$ are the smaller and greater of the roots of

$$\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2r\frac{x-\bar{x}}{\sigma_x}\frac{vx-\bar{y}}{\sigma_y} + \left(\frac{vx-\bar{y}}{\sigma_y}\right)^2 = \lambda^2 \quad \text{..........(30)}.$$

Writing

$$\left.\begin{aligned}
\alpha &= \frac{\text{[illegible]}}{\sigma_x^2\sigma_y^2} \\
\beta &= -\frac{\sigma_y(r\bar{y}\sigma_x - \bar{x}\sigma_y) + v\sigma_x(r\bar{x}\sigma_y - \bar{y}\sigma_x)}{\sigma_x^2\sigma_y^2} \\
\gamma &= \frac{\bar{x}^2}{\sigma_x^2} - 2r\frac{\bar{x}}{\sigma_x}\frac{\bar{y}}{\sigma_y} + \frac{\bar{y}^2}{\sigma_y^2} \\
\epsilon &= \gamma - \frac{\beta^2}{\alpha} = \frac{(1-r^2)(v\bar{x}-\bar{y})^2}{\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2}
\end{aligned}\right\} \quad \text{..........(31)},$$

we have, for (29) and (30),

$$I(\lambda) = z_0 \int_{x_1}^{x_2} |x| \, e^{-\frac{1}{2}\frac{1}{1-r^2}(\alpha x^2 - 2\beta x + \gamma)} dx \quad \text{..........(32)},$$

where x_1 and x_2 are the roots of

$$\alpha x^2 - 2\beta x + \gamma = \lambda^2,$$

i.e.

$$\alpha\left(x - \frac{\beta}{\alpha}\right)^2 = \lambda^2 - \epsilon \quad \text{..........(33)}.$$

Now if the ellipse (28) lies entirely in the positive quadrant, the index must lie between two positive values given by the gradients of the tangents to the ellipse from the origin. For these values x_1 and x_2 coincide, so that they are given by

$$\lambda^2 - \epsilon = 0.$$

For all values of v lying between these limits, x_1 and x_2 are different and positive so that $\beta > 0$, and (32) gives

$$I(\lambda) = z_0 \int_{x_1}^{x_2} x e^{-\frac{1}{2}\frac{1}{1-r^2}(\alpha x^2 - 2\beta x + \gamma)} dx$$

$$= z_0 e^{-\frac{1}{2}\frac{\epsilon}{1-r^2}} \left\{\int_{x_1}^{x_2}\left(x - \frac{\beta}{\alpha}\right) e^{-\frac{1}{2}\frac{\alpha}{1-r^2}\left(x-\frac{\beta}{\alpha}\right)^2} dx + \int_{x_1}^{x_2} \frac{\beta}{\alpha} e^{-\frac{1}{2}\frac{\alpha}{1-r^2}\left(x-\frac{\beta}{\alpha}\right)^2} dx\right\}.$$

The first of these integrals vanishes, since x_1 and x_2 are the roots of (33); putting, in the second,

$$\frac{\alpha}{1-r^2}\left(x-\frac{\beta}{\alpha}\right)^2=t^2,$$

we find $$I(\lambda)=\sqrt{2\pi}\,\frac{z_0\beta\sqrt{1-r^2}}{\alpha^{\frac{3}{2}}}\,e^{-\frac{1}{2}\frac{\epsilon}{1-r^2}}\int_{\left(x_1-\frac{\beta}{\alpha}\right)\sqrt{\frac{\alpha}{1-r^2}}}^{\left(x_2-\frac{\beta}{\alpha}\right)\sqrt{\frac{\alpha}{1-r^2}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}\,dt,$$

or, by (33),

$$I(\lambda)=c_0\frac{\beta}{\alpha^{\frac{3}{2}}}e^{-\frac{1}{2}\frac{\epsilon}{1-r^2}}\int_0^{\sqrt{\frac{\lambda^2-\epsilon}{1-r^2}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}\,dt \qquad\text{.........................(34)},$$

where c_0 is to be determined so as to make the area under the frequency curve of v equal to unity.

Application to Anthropometric Data.

We return to equation (17). If $\bar{x}$ be large compared with σ_x, h will be large; accordingly, $\int_h^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}X^2}dX$ will be negligible and so, *a fortiori*, will

$$\int_h^\infty\int_k^\infty \frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{1-\rho^2}(X^2-2\rho XY+Y^2)}\,dX\,dY$$

and $$\int_{-\infty}^{-h}\int_{-k}^\infty \frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{1-\rho^2}(X^2-2\rho XY+Y^2)}\,dX\,dY.$$

Hence the chance C of an index not less than v will be, approximately,

$$C\simeq\int_{-k}^\infty\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Y^2}dY=\int_{\frac{v\bar{x}-\bar{y}}{(\sigma_y^2-2\cdot v\sigma_x\sigma_y+v^2\sigma_x^2)^{\frac{1}{2}}}}^\infty\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Y^2}\,dY\qquad\text{.........(35)}.$$

Thus we have Geary's result*, that if the ratio $\bar{x}/\sigma_x$ be large, then, approximately,

$$\frac{v\bar{x}-\bar{y}}{(\sigma_y^2-2rv\sigma_x\sigma_y+v^2\sigma_x^2)^{\frac{1}{2}}}$$

is distributed normally with unit standard deviation.

Differentiating (35), we see that the equation to the index-distribution will be, approximately,

$$\psi(v)=-\frac{\partial C}{\partial v}\simeq\frac{\sigma_y(\bar{x}\sigma_y-r\bar{y}\sigma_x)+v\sigma_x(\bar{y}\sigma_x-r\bar{x}\sigma_y)}{(\sigma_y^2-2rv\sigma_x\sigma_y+v^2\sigma_x^2)^{\frac{3}{2}}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(v\bar{x}-\bar{y})^2}{\sigma_y^2-2rv\sigma_x\sigma_y+v^2\sigma_x^2}}$$
$$\text{.........(36)},$$

a form that can be deduced from (24) by neglecting the first term in the second member, and replacing the upper limit of the integral factor in the second term by $-\infty$.

* R. C. Geary: "The Frequency Distribution of the Quotient of Two Normal Variates," *Journal of the Royal Statistical Society*, Vol. XCIII. 1930, p. 442.

It is worth emphasising the conditions under which this approximation is valid. It is easy to see that for some values of v the ordinate calculated from (36) will be negative; but if $\int_{\bar{x}/\sigma_x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$ vanishes to r decimal places, then in numerical work proceeding to r decimal places or less, equations (17) and (35) will appear to supply exactly the same values for the frequencies; in other words, the negative frequencies furnished by (35) will be zero, to the degree of accuracy of our calculations.

Now let us return to equation (34); it shows that the effect of neglecting the values of x, y that lie outside the ellipse (28) is to change the distribution of

$$u = \frac{v\bar{x} - \bar{y}}{\sqrt{\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2}} = \sqrt{\frac{\epsilon}{1 - r^2}}$$

from a form that is sensibly normal to the form

$$z = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} c_0' \int_0^{\sqrt{\frac{\lambda^2}{1-r^2} - u^2}} e^{-\frac{1}{2}t^2} dt \qquad \text{.....................(37).}$$

The ordinate of the normal distribution is thus multiplied by a factor that decreases with the length of the ordinate; the effect on the appearance of the curve of distribution of u, and therefore on that of the curve of distribution of v, will be to increase the areas near the mode at the expense of the tails. But this effect may easily be invisible, if λ is at all large; if, for example, λ^2 be eighteen times $1 - r^2$, no ordinate of the normal curve within a range $\pm 3\sigma$ $(\sigma = 1)$ of the mean will be altered by much more than one-tenth per cent.

Thus we have the somewhat startling result, that if $\bar{x}$ and $\bar{y}$ be positive, and large compared with their standard deviations, then the limitation of x and y to finite positive values can change the moments of the index-distribution from infinite to finite values, without having any visible effect on the appearance of the distribution. The paradox disappears, when we consider the difference between mathematical formulae and their numerical representation. Infinite moments cannot occur in the numerical applications of mathematical theory, any more than they can occur in experimental sampling. Any calculation of frequencies from the mathematical equation to a distribution will be performed to a certain number of decimal places; as calculated from the numerical frequencies, the moments will always be finite. When we say that the distribution has infinite moments, we mean that the numerical values of the moments do not tend to any finite limits, as we increase indefinitely the number of places to which we work; but this remark has no practical interest—it is, in fact, irrelevant.

What the computer about to fit a curve wants to know, are the values that the moments would have, for the set of numerical frequencies calculated to the number of decimal places that he intends working to. In any discussion of anthropological data*, this number will be small enough to justify the use of (35) in place of (17),

[* This appears to exclude from anthropological data such a character as corneal astigmatism, where the mean = ·62 dioptres, and the standard deviation ·86 dioptres; thus the character may be negative as well as positive. An index formed by the ratio of corneal astigmatism to distance of near point would need (24) rather than (36). ED.]

and the frequencies will appear to be those under (36), or, if λ be taken large enough, under (34). But since (34) is a distribution of limited range, the numerical values of its moments, as calculated from the numerical frequencies, will not differ appreciably from the values obtained by a direct mathematical process. This, I think, rather than any *a priori* rejection, as impossible, of zero or infinite values of the variates, is the true justification of Merrill's method of approaching the subject*.

If ξ and η be the deviations of x and y from their mean values, so that

$$x = \bar{x} + \xi, \quad y = \bar{y} + \eta,$$

Merrill writes
$$v = \frac{\bar{y} + \eta}{\bar{x} + \xi} = \frac{\bar{y}}{\bar{x}}\left(1 + \frac{\eta}{\bar{y}}\right)\left(1 - \frac{\xi}{\bar{x}} + \frac{\xi^2}{\bar{x}^2} - \frac{\xi^3}{\bar{x}^3} + \dots\right) \quad \dots\dots\dots\dots\dots(38),$$

and takes for μ_n', the nth moment about zero of the distribution of v, the mean value of

$$v^n = \left(\frac{\bar{y}}{\bar{x}}\right)^n\left(1 + \frac{\eta}{\bar{y}}\right)^n\left(1 - \frac{\xi}{\bar{x}} + \frac{\xi^2}{\bar{x}^2} - \frac{\xi^3}{\bar{x}^3} + \dots\right)^n \quad \dots\dots\dots\dots\dots(39);$$

to obtain this value Merrill retains the products $\xi^r\eta^s$ as far as the eighth order and takes for their mean values the product moments p_{rs} of the normal surface (6).

This process would not be valid if we imagined it applied to the whole of the xy-plane, since the expansion (38) holds only if $|\xi| < \bar{x}$; but it is valid, if applied to the interior of any probability contour (28) that lies in the positive quadrant. In the case of low-order product-moments, we do not commit any serious error in taking for the product-moments of the curtailed distribution (27) the values derived from the whole normal surface (6), so that Merrill's values of the moments may legitimately be regarded as the moments of the index-distribution in a curtailed normal population, which are exactly what we want.

Illustration.

I have illustrated the preceding theory on some figures kindly supplied by Dr T. L. Woo from his measurements on the Biometric Laboratory's series of Egyptian skulls. Table I shows the joint distribution of $T_1(L)$ and $P_2(L)$†, two measurements made, on the temporal and parietal bones respectively, in the left-hand side of 787 skulls.

Taking $P_2(L)$ as x, $T_1(L)$ as y, we have‡

$$\left.\begin{array}{ll} \bar{x} = 111{\cdot}207\,433, & \bar{y} = 86{\cdot}019\,060 \\ \sigma_x = 5{\cdot}7885^-, & \sigma_y = 3{\cdot}8453 \\ & r_{xy} = {\cdot}173833 \end{array}\right\} \quad \dots\dots\dots\dots\dots\dots\dots(40).$$

* A. S. Merrill, "Frequency Distribution of an Index when Both the Components Follow the Normal Law," *Biometrika*, Vol. xx^A. 1928, pp. 53—63.

† For the precise definition of these measurements see T. L. Woo, "On the Asymmetry of the Human Skull," *Biometrika*, Vol. xxii. 1930—31, pp. 326 and 327.

‡ These constants, and those of the sample distribution of the index, were kindly calculated by Dr T. L. Woo.

TABLE I. *Distribution of $T_1(L)$ and $P_2(L)$.*

$T_1(L)$ (Central Values)

$P_2(L)$ (Central Values)	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	Totals
89·75	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
90·75	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
91·75	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
92·75	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
93·75	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
94·75	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
95·75	—	—	—	—	1	1	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	3
96·75	1	—	—	—	1	—	—	—	—	—	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	4
97·75	—	—	—	1	—	1	—	—	—	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	4
98·75	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	2
99·75	—	—	1	—	—	—	—	—	—	—	2	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	4
100·75	—	—	—	—	—	1	—	—	—	2	2	—	1	1	—	—	—	—	—	1	—	1	—	—	—	—	9
101·75	—	—	—	1	—	—	1	—	1	—	—	3	—	—	2	1	1	—	—	—	—	—	—	—	—	—	10
102·75	—	—	—	—	1	—	2	2	—	1	4	1	2	1	—	—	1	—	—	—	1	—	—	—	—	—	16
103·75	—	1	—	—	—	—	1	—	1	4	3	1	1	—	2	1	1	—	1	—	—	—	—	—	—	—	17
104·75	—	—	—	1	—	—	1	2	—	4	2	4	2	3	1	4	3	—	—	—	—	—	—	—	—	—	27
105·75	—	—	—	—	—	1	2	4	3	—	3	5	6	2	2	2	2	3	1	2	—	2	1	—	—	1	42
106·75	—	—	1	—	—	1	—	1	2	4	6	5	4	3	4	5	3	3	—	—	—	—	—	—	—	—	42
107·75	—	—	—	1	1	1	2	2	4	5	4	6	11	7	3	4	2	1	1	—	—	—	—	—	—	—	55
108·75	—	—	—	1	1	—	1	1	6	4	7	6	8	6	6	1	5	1	—	—	2	—	1	—	—	—	57
109·75	—	—	—	—	1	1	—	7	6	4	7	1	3	7	7	5	4	1	3	1	2	—	—	—	—	1	61
110·75	—	—	—	—	—	3	3	1	4	7	6	5	7	2	8	5	2	3	4	—	1	—	—	—	—	—	61
111·75	—	—	—	—	1	1	1	1	3	2	5	7	5	3	4	7	5	1	3	2	—	—	—	—	—	—	51
112·75	—	1	—	—	—	—	—	—	4	4	2	5	8	5	3	3	2	1	1	—	1	—	—	—	—	—	40
113·75	—	—	—	—	—	—	1	2	3	5	2	6	7	6	2	2	4	1	1	1	—	2	—	—	—	—	45
114·75	—	—	—	—	—	1	1	2	3	4	4	7	5	5	1	3	2	2	3	1	1	—	—	—	—	—	45
115·75	—	—	—	—	—	—	—	2	4	3	5	3	4	6	3	5	1	5	—	—	—	—	—	—	—	—	41
116·75	—	—	—	—	—	—	2	—	1	2	6	1	4	3	5	2	3	2	1	—	—	—	—	—	—	—	32
117·75	—	—	—	—	1	1	—	3	1	—	2	—	5	4	1	2	4	3	—	—	—	1	—	—	—	—	28
118·75	—	—	—	—	1	—	—	1	—	1	—	2	2	2	1	2	3	2	2	1	—	—	—	—	—	—	20
119·75	—	—	—	—	—	2	—	—	—	2	2	2	1	1	—	2	2	—	—	1	—	—	—	2	—	—	17
120·75	—	—	—	—	—	—	—	1	3	—	1	1	3	2	—	—	1	2	1	1	—	—	—	—	—	—	16
121·75	—	—	—	—	—	—	—	—	—	—	—	2	2	1	2	—	—	—	—	—	1	—	1	—	—	—	9
122·75	—	—	—	—	—	—	—	—	—	—	—	1	—	1	2	1	1	1	1	—	—	1	—	—	—	—	9
123·75	—	—	—	—	—	—	—	—	1	2	—	1	—	—	1	—	—	—	1	—	—	—	—	—	—	—	6
124·75	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	3
125·75	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	2
126·75	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1
127·75	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
128·75	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
129·75	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
130·75	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
131·75	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
132·75	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
133·75	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
Totals	1	2	2	6	10	15	18	33	51	61	77	78	91	74	63	60	52	32	27	11	9	7	3	2	—	2	787

TABLE II. *The Index Distribution.*

Frequencies			Ordinates		
Central Values	Observed	Calculated	v	From (41)	From (42)
·60	—	·01	·595	1	1
·61	—	·06	·605	3	3
·62	1	·14	·615	9	9
·63	—	·34	·625	22	22
·64	—	·75	·635	50	50
·65	1	1·54	·645	106	106
·66	6	2·93	·655	211	211
·67	6	5·21	·665	389	389
·68	6	8·67	·675	672	671
·69	12	13·56	·685	1086	1086
·70	16	19·93	·695	1650	1649
·71	31	27·64	·705	2359	2359
·72	30	36·22	·715	3184	3184
·73	49	44·98	·725	4064	4065
·74	47	53·03	·735	4920	4921
·75	69	59·49	·745	5660	5661
·76	77	63·65	·755	6200	6202
·77	63	65·05	·765	6483	6484
·78	56	63·70	·775	6483	6483
·79	60	59·80	·785	6212	6211
·80	67	53·96	·795	5715	5714
·81	46	46·89	·805	5057	5056
·82	33	39·30	·815	4313	4312
·83	33	31·83	·825	3550	3550
·84	27	24·95	·835	2826	2826
·85	13	18·96	·845	2179	2179
·86	13	13·99	·855	1630	1630
·87	6	10·03	·865	1185	1185
·88	4	6·99	·875	838	838
·89	2	4·79	·885	578	577
·90	4	3·17	·895	388	388
·91	3	2·06	·905	255	255
·92	2	1·31	·915	164	164
·93	1	·82	·925	103	103
·94	2	·50	·935	64	64
·95	2	·30	·945	38	38
·96	—	·18	·955	23	23
·97	—	·10	·965	13	13
·98	—	·06	·975	8	8
·99	—	·03	·985	4	4
1·00	—	·02	·995	2	2
1·01	—	·01	1·005	1	1

The distribution of the index $v = T_1(L)/P_2(L)$ in the 787 skulls is shown in the first column of Table II.

Let us assume that $P_2(L)$ and $T_1(L)$ are distributed in a normal surface whose constants are given by (40)*. We have $\bar{x}/\sigma_x = 19{\cdot}22$, so that $\int_{\bar{x}/\sigma_x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx$

* Actually we have for $P_2(L)$: $\beta_1 = {\cdot}0002$, $\beta_2 = 3{\cdot}5340$,
for $T_1(L)$: $\beta_1 = {\cdot}0064$, $\beta_2 = 3{\cdot}1576$.

From E. S. Pearson's table of the 5 °/₀ and 1 °/₀ points of the distribution of β_1 and β_2 (*Biometrika*, Vol. XXII. 1930—31, p. 248; or *Tables for Statisticians and Biometricians*, Vol. II. Table XXXVII *bis*),

vanishes to something like 80 places of decimals. The frequencies of the theoretical index-distribution (24) may therefore be calculated from (35); they are given in the second column of Table II. (Their calculation can be effected very rapidly, by first forming a column of the values of $\dfrac{v\bar{x}-\bar{y}}{(\sigma_y^2 - 2rv\sigma_x\sigma_y + \sigma_x^2)^{\frac{1}{2}}}$ corresponding to the boundaries of the frequency groups, and then interpolating into tables of the normal curve.) Combining the tails of the theoretical distribution, above ·895 and below ·675, into two single classes, and grouping together the frequencies whose centres are ·88 and ·89, we find $\chi^2 = 25{\cdot}1504$ and $P = {\cdot}290$, so that it is quite likely that the sample shown in the first column is a random sample from the parent population shown in the second column in Table II.

In Table III are shown the constants of the index-distribution, calculated

(i) from the observed distribution,

(ii) from the numerical frequencies given for the theoretical distribution,

(iii) from Merrill's formulae*.

TABLE III. *Frequency Constants of the Index-Distribution.*

Frequency Constant	Calculated from		
	(i) Col. 1, Table II	(ii) Col. 2, Table II	(iii) Merrill's formulae
Mean	·77469	·775296	·775298
S. D.	·048280	·048646	·048664
β_1	·1111	·0506	0511
β_2	3·6296	3·1073	3·1196

It will be observed that the agreement between corresponding entries in the last two columns is quite satisfactory.

If we substitute in equation (24) the values of $\bar{x}$, $\bar{y}$, σ_x, σ_y, and r given by (40), we find that throughout the effective range of the distribution of v, the first term in $\psi(v)$ is less than e^{-360}, while the upper limit of the integral factor in the second term is in the neighbourhood of -27. These figures indicate the extreme accuracy with which the index-distribution is represented by (36). Substituting in that equation, and multiplying the second member by 787, the size of the sample, we find for the equation to the index-distribution

$$y = \frac{1032191 + 192966v}{(14{\cdot}7867 - 7{\cdot}7386v + 33\,5067\,v^2)^{\frac{3}{2}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(86{\cdot}01906 - 111{\cdot}20743v)^2}{14{\cdot}7867 - 7{\cdot}7386v + 33{\cdot}5067v^2}} \quad \text{.........(41);}$$

the Pearson Type IV curve, fitted from Merrill's formulae for the moments, is

$$y = 10^{-20} \times 7{\cdot}83548\,(1 + 5{\cdot}36951x^2)^{-73{\cdot}28151}\,e^{131{\cdot}34936\tan^{-1}2{\cdot}31722x} \quad \text{...(42).}$$

we find that while the values of β_1 are not significant of any departure from normality, β_2 exceeds 3·30 in less than 5 °/₀ and 3·48 in less than 1 °/₀ of random samples of 787 from a normal population. It is accordingly very unlikely that $P_2(L)$ is distributed normally, but it will be seen that the departure from normality does not seriously affect the goodness of fit of the index-distribution.

* *Loc. cit.* pp. 53, 56, 57.

The ordinates of these two curves are shown in the last two columns of Table II; the agreement between them is practically exact. The present example would appear to indicate, therefore, that there will in practice be little difference between the conclusions suggested by the two theories. Nevertheless, once it has been assumed that the components of the index are normally distributed, it seems definitely preferable to use the index-distribution (24), and its probability integral (17), implied in this assumption of normality. We thereby retain a consistent mathematical theory, by avoiding the extraneous assumption that the index-distribution can be represented by one of the Pearsonian curves; moreover, calculating the ordinates and frequencies for Geary's approximation to (24) is considerably less laborious than doing so for a Type IV curve.

It will be observed that while we get a fairly satisfactory fit of the theoretical to the observed index-distribution, the latter is more leptokurtic than the former. It is therefore to be expected that the fit will be improved, if we assume that $T_1(L)$ and $P_2(L)$ are distributed in a curtailed normal surface such as (27), so that the distribution of the index becomes of the type (34). Actually, however, we find that for the outlying individual for which $T_1(L) = 92$, $P_2(L) = 134$,

$$\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2r\frac{x-\bar{x}}{\sigma_x}\frac{y-\bar{y}}{\sigma_y} + \left(\frac{y-\bar{y}}{\sigma_y}\right)^2 = 15{\cdot}7947,$$

so that unless we are to reject this individual as a pathological anomaly, we must take the λ^2 of the limiting contour (28) at least as large as 16. The following table indicates the sort of value that F, the integral factor in $I(\lambda)$ (equation 34), assumes, if we adopt this value for λ^2:

v:	·625	·635	·645	·655	·695	·735	·775
$2F$:	·961	·986	·994	·997	·9998	·99993	·99995
v:	·955	·945	·935	·925	·885	·845	·805
$2F$:	·980	·989	·994	·996	·9994	·99986	·99994

These values make it clear that there will not be any significant difference between the ordinates deduced from (34) and those shown in Table II; in other words, there will be no appreciable improvement in the fit.

I have to thank Professor Pearson for his advice at several points in the course of this paper, and for lending me the manuscript of an unpublished lecture by him on Dr Merrill's work*.

* [The purpose of the lecture referred to lay in pointing out from a number of examples, whose components were even more nearly normal than those of Woo's case, that the β_2's of index-distributions were very considerably in excess of their theoretical value as deduced from Merrill's process. Mr Fieller shows that Merrill's method and his own lead to results in fair accordance both in the distribution of frequency and in the values of the constants. Assuming the theoretical values of both methods to give $\beta_1 = {\cdot}10$ and $\beta_2 = 3{\cdot}1$, the standard error of β_2 for a sample size 787 is ·2395. Thus the observed β_2 has a deviation from the theoretical value of 2·18 times its standard error, and roughly this would indicate a probability of less than ·02 of the observed β_2 being due to a random sampling from the theoretical population. Thus the difficulty discussed in the lecture is emphasised rather than surmounted by a method which gives constants in accordance with Merrill's results. Ed.]

A NOTE ON THE DISTRIBUTION OF THE CORRELATION RATIO.

By JOHN WISHART, M.A., D.Sc., Clare College, Cambridge.

Introduction.

THE sampling distribution of the correlation ratio may now be said to have been determined for three different cases, in all of which the arrays of a variable, y, are normally distributed with a common variance in the population sampled. Suppose that we reserve the symbol η^2 for the population value and denote by E^2 the square of the correlation ratio calculated from a sample. That is

$$E^2 = \sum_1^p \{n_i(\bar{y}_i - \bar{y})^2\}/\{\Sigma(y-\bar{y})^2\} \quad \text{.........................(1)}.$$

The number of arrays is p, the ith array having n_i observations; $\bar{y}_i$ is the mean of the observations in the ith array, and $\bar{y}$ is the general mean. The first summation is over all arrays, while the second is for all the observations of the sample, i.e. from 1 to N, where $N = \sum_1^p (n_i)$.

Then it is known that *if* η^2 *be zero**, the distribution of E^2 is

$$df = \frac{\{\frac{1}{2}(n_1+n_2-2)\}!}{\{\frac{1}{2}(n_1-2)\}!\,\{\frac{1}{2}(n_2-2)\}!}(E^2)^{\frac{1}{2}(n_1-2)}(1-E^2)^{\frac{1}{2}(n_2-2)}\,d(E^2) \quad \text{......(2)},$$

a form which shows the identity of the form with a general class of distributions having symmetry in n_1 and n_2, interpreted in this case as the numbers of degrees of freedom between and within arrays, so that

$$n_1 = p-1, \quad n_2 = N-p.$$

If η^2 *be not zero*, two separate cases arise, for both of which solutions can be derived from the two distributions of the multiple correlation coefficient given by R. A. Fisher†.

Case (a). Here it is to be supposed that the conditions of sampling are such that the array totals, n_i, vary from sample to sample. The sampling distribution is then given by Fisher's series (A), writing

$$R^2 = E^2, \quad \rho^2 = \eta^2, \quad n_1 = p-1, \quad n_2 = N-p,$$

provided that the expectations of y for the values of x in the sampled population are normally distributed. This distribution has been studied at some length by Fisher in the paper cited, and in particular the probability integral for n_2 even

* R. A. Fisher: *Journ. Roy. Stat. Soc.* Vol. LXXXV. 1922, p. 605. The distribution was also deduced at a later date by Hotelling: *Proc. Nat. Acad. Sc.* Vol. XI. 1925, pp. 657—662.

† R. A. Fisher: *Proc. Roy. Soc.* A, 121, 1928, pp. 654—673.

was expressed in finite terms. As it was thought that the mean and second moment coefficient of this distribution were of some interest in themselves, these quantities were later determined for n_2 even, and the results inferred to hold generally*. Such results can readily be translated into terms of the correlation ratio.

Case (*b*). Of more practical interest, however, is the case where the number n_i in each array is supposed the same for all samples. The distribution of E^2 is then that of R^2 in Fisher's distribution (C), writing

$$R^2 = E^2, \quad \tfrac{1}{2}\beta^2 = \frac{N}{2}\frac{\eta^2}{1-\eta^2}, \quad n_1 = p-1, \quad n_2 = N-p.$$

Fisher did not study the properties of distribution (C), and the object of the present paper is to obtain the probability integral in finite terms, and in addition expressions for the mean and second moment coefficients, developing the distribution as that of E^2, the square of the correlation ratio, although the results can readily be translated in terms of any variate following the same law of distribution.

Further, both Fisher's (A) and (C) distributions were shown to tend in the limit as the size of the sample was increased to a third distribution (B), and it will be shown how the results derived in this paper, equally with those previously obtained from the (A) distribution, tend in the limit to the corresponding parameters of the distribution (B). For this limiting form we have

$$\beta^2 = n_2\eta^2, \quad B^2 = n_2 E^2.$$

The (C) *Distribution.*

Changing the notation as explained above, and denoting $\frac{1}{2}n_1$ by a, $\frac{1}{2}n_2$ by b, E^2 by x and $\frac{1}{2}\beta^2$ by t to simplify the mathematics, the distribution of x takes the form

$$df = e^{-t}\frac{(a+b-1)!}{(a-1)!\,(b-1)!}\,x^{a-1}(1-x)^{b-1}$$
$$\left[1 + \frac{a+b}{1!\,a}(tx) + \frac{(a+b)(a+b+1)}{2!\,a(a+1)}(tx)^2 + \ldots\right]dx \quad \ldots\ldots(3).$$

Since n_1 and n_2 are necessarily whole numbers, a and b may be integers or half integers, but the factorial sign is used in either case, i.e. $x!$ denotes what is generally understood by $\Gamma(x+1)$. The series within square brackets is a confluent hypergeometric one, and may be denoted by $F(a+b, a, tx)$. Now by an application of Kummer's formula†, we find

$$F(a+b, a, tx) = e^{tx}F(-b, a, -tx),$$

giving, when b is an integer, a terminating series of the form

$$1 + \frac{b}{1!\,a}(tx) + \frac{b(b-1)}{2!\,a(a+1)}(tx)^2 + \ldots.$$

* J. Wishart: *Biometrika*, Vol. XXII. 1931, pp. 358—361.

† Whittaker and Watson: *Modern Analysis*, § 16.11 (2nd edn. p.

At this point we shall suppose b to be an integer, returning later to a consideration of the other case, when b is a half-integer. Noting that

$$\frac{d^b}{dx^b}(x^{a+b-1}e^{tx}) = \frac{(a+b-1)!}{(a-1)!}x^{a-1}e^{tx}F(-b, a, -tx),$$

the distribution (3) may be written in the comparatively simple form

$$df = \frac{e^{-t}}{(b-1)!}(1-x)^{b-1}f^{(b)}(x)\,dx \quad \text{......................(4)},$$

where $f(x) = x^{a+b-1}e^{tx}$, and $f^{(b)}(x)$ denotes the bth differential coefficient of $f(x)$ with respect to x.

Probability Integral.

In this form it is easy to evaluate the indefinite or probability integral of the distribution. The range in (4) is from 0 to 1, and since $f^{(r)}(0) = 0$ for all values of r from 0 to $b-1$, the integral of (4) from 0 to x may be written down directly. In fact,

$$\begin{aligned} I &= \int_0^x df \\ &= e^{-t}\sum_{r=0}^{b-1}\frac{(1-x)^r}{r!}f^{(r)}(x) \\ &= \frac{(a+b-1)!}{e^{t(1-x)}}\sum_{r=0}^{b-1}\left\{\frac{x^{a+b-1-r}(1-x)^r}{r!\,(a+b-1-r)!}F(-r, a+b-r, -tx)\right\} \quad \text{......(5)}, \end{aligned}$$

involving a series which terminates in $\frac{1}{2}b(b+1)$ elementary terms.

Now by Taylor's theorem

$$f(x+h) = \sum_{r=0}^{\infty}\frac{h^r}{r!}f^{(r)}(x).$$

Put $h = 1-x$, and we see that (5) involves the first b terms of an expansion in Taylor's series, of which the complete series is

$$f(x+1-x) = f(1) = e^t.$$

When, therefore, we are interested in the "tail" of the distribution curve, as when we wish to find a value of x for which the proportionate area under the curve beyond the ordinate at x is 0·02 or 0·01, say, we may write the probability integral in the form

$$I = 1 - e^{-t}\sum_{r=b}^{\infty}\frac{(1-x)^r}{r!}f^{(r)}(x) \quad \text{.....................(6)}.$$

If it were desired to extend Woo's table* for values of η^2 other than zero, it would be necessary to solve equations for x of the form

$$\sum_{r=b}^{\infty}\frac{(1-x)^r}{r!}g^{(r)}(x) = 0{\cdot}02 \text{ and } 0{\cdot}01,$$

where
$$g^{(r)}(x) = e^{-t}f^{(r)}(x) = \frac{d^r}{dx^r}(x^{a+b-1}e^{-t(1-x)}).$$

* T. L. Woo: *Biometrika*, Vol. XXI. 1929, pp. 1—66. *Tables for Statisticians and Biometricians*, Part II, 1931, pp. 16—72. A table of the author's, in *Quart. Journ. Roy. Met. Soc.* Vol. LIV. 1928, pp. 258—259, gives the 0·05 and 0·01 levels of significance, and extends to 7 arrays and a size of sample of about 100. It covers a range below 50, which is not covered by Woo's table.

For given size of sample and number of arrays, and for a given value of η^2, this would give an x (or E^2) beyond which there is only a 1 in 50 or a 1 in 100 chance of a value occurring in samples from a population with this value of η^2. At best, however, this computation will be a long and laborious business, for it does not appear that any essential simplification is possible in the expression of our results (5) and (6).

For the special case of samples from uncorrelated data, Woo's tables provide, in addition to the approximate 0·01 and 0·02 probability levels of significance, the mean value and standard deviation of the square of the sample correlation ratio (our x or E^2). We shall proceed now to determine these quantities in the general case, still on the assumption that b is an integer.

Mean value of E^2.

Beginning with the form (4) of the distribution, let us multiply by x and integrate from 0 to 1. We have

$$\overline{E^2} = \int_0^1 \frac{e^{-t}}{(b-1)!} x(1-x)^{b-1} d\{f^{(b-1)}(x)\}$$

$$= \left[\frac{e^{-t}}{(b-1)!} x(1-x)^{b-1} f^{(b-1)}(x)\right]_0^1 - \int_0^1 \frac{e^{-t}}{(b-1)!}\{(1-x)^{b-1} - (b-1)x(1-x)^{b-2}\} f^{(b-1)}(x)\,dx,$$

on integrating by parts,

$$-\int_0^1 \frac{e^{-t}}{(b-1)!}\{(1-x)^{b-1} - (b-1)x(1-x)^{b-2}\} f^{(b-1)}(x)\,dx,$$

since the term between limits vanishes. Continuing this process we have finally

$$\overline{E^2} = \int_0^1 \frac{e^{-t}}{(b-1)!}\{-(b-1)(b-1)!(1-x) + (b-1)!\,x\} f'(x)\,dx$$

$$= e^{-t}\int_0^1 \{1 - b(1-x)\}\, d(x^{a+b-1} e^{tx})$$

$$= 1 - be^{-t}\int_0^1 x^{a+b-1} e^{tx}\,dx \quad \text{.....................(7)}$$

on further integration by parts. An examination of the integral in (7) shows that when $t = 0$ we have

$$\overline{E^2} = 1 - \frac{b}{a+b} = \frac{a}{a+b} = \frac{n_1}{n_1+n_2} = \frac{p-1}{N-1},$$

agreeing with the result deduced directly from the special form (2) of the distribution when $\eta^2 = 0$. On the other hand when t becomes very large the important part of the integrand is e^{tx}, whose integral from 0 to 1 is $(e^t - 1)/t$, so that the second member of (7) behaves like $b(1-e^{-t})/t$, which tends to zero as t tends to infinity. Thus when $\eta^2 = 1$, we have $\overline{E^2} = 1$.

The integral in (7) may be evaluated in series form in two ways, according to

whether t is less or greater than $a+b$. In the first case, expanding the exponential and integrating term by term, we have

$$\int_0^1 x^{a+b-1} e^{tx}\,dx = \left[\frac{x^{a+b}}{a+b} + t\frac{x^{a+b+1}}{a+b+1} + \frac{t^2}{2!}\frac{x^{a+b+2}}{a+b+2} + \dots\right]_0^1$$

$$= \frac{1}{a+b} F(a+b,\, a+b+1,\, t)$$

$$= \frac{e^t}{a+b} F(1,\, a+b+1,\, -t)$$

by Kummer's formula, F being the confluent hypergeometric series as already defined. Thus, in general,

$$\bar{E}^2 = 1 - \frac{b}{a+b} F(1,\, a+b+1,\, -t) \quad \dots\dots\dots\dots(8).$$

The series F is of the form

$$\frac{t}{a+b+1} + \frac{t^2}{(a+b+1)(a+b+2)} - \dots,$$

and can be readily evaluated when $a+b+1$ is large compared with t. When t is large, however, or when $a+b$ is small whatever the value of t be, it will be found best to transform (8), a process which is best carried out by returning to the integral in (7) and integrating by parts. We have:

$$\frac{1}{t}\int_0^1 x^{a+b-1} d(e^{tx}) = \frac{e^t}{t} - \frac{a+b-1}{t}\int_0^1 x^{a+b-2} e^{tx}\,dx$$

$$= \frac{e^t}{t}\left[1 - \frac{a+b-1}{t} + \frac{(a+b-1)(a+b-2)}{t^2} - \dots + \frac{(-1)^{a+b-1}(a+b-1)!}{t^{a+b-1}}(1-e^{-t})\right]$$

when $(a+b)$ is an integer. When $(a+b)$ is a half-integer, the other possibility, the corresponding form is

$$\frac{e^t}{t}\left[1 - \frac{a+b-1}{t} + \frac{(a+b-1)(a+b-2)}{t^2} - \dots + \frac{(-1)^{a+b-\frac{3}{2}}(a+b-1)\dots\frac{3}{2}}{t^{a+b-\frac{3}{2}}}\right]$$
$$+ \frac{(-1)^{a+b-\frac{1}{2}}(a+b-1)\dots\frac{1}{2}}{t^{a+b-\frac{1}{2}}}\int_0^1 x^{-\frac{1}{2}} e^{tx}\,dx.$$

Putting $x = u^2$ in the final integral, the series becomes

$$\frac{e^t}{t}\left[1 - \frac{a+b-1}{t} + \frac{(a+b-1)(a+b-2)}{t^2} - \dots\right.$$
$$\left. + \frac{(-1)^{a+b-\frac{3}{2}}(a+b-1)\dots\frac{3}{2}}{t^{a+b-\frac{3}{2}}}\left(1 - e^{-t}\int_0^1 e^{tu^2}\,du\right)\right].$$

We therefore have

$$E^2 = 1 - \frac{b}{t}\left[1 - \frac{a+b-1}{t} + \frac{(a+b-1)(a+b-2)}{t^2}\right.$$
$$\left. + \frac{(-1)^{a+b-1}(a+b-1)!}{t^{a+b-1}}(1-e^{-t})\right] \; (a+b) \text{ an integer} \quad \dots\dots(9),$$

$$+ \frac{(-1)^{a+b-\frac{3}{2}}(a+b-1)\dots\frac{3}{2}}{t^{a+b-\frac{3}{2}}}\left(1 - e^{-t}\int_0^1 e^{tu^2}\,du\right)\Bigg] \; (a+b) \text{ a half-integer} \;\dots(10).$$

The series in (9) and (10) consist of a finite number of terms, from which the value of $\overline{E}^2$ may be readily computed. When t is large compared with $(a+b)$, it may happen that the terms become negligible before the last, or remainder term is reached. If, however, this last term has to be computed, there is no difficulty with (9), but with (10) we have to consider methods of evaluating the integral. Let $t(1-u^2)=x$. Then

$$\int_0^1 e^{-t(1-u^2)}du = \frac{1}{2t}\int_0^t \frac{e^{-x}dx}{\sqrt{1-\frac{x}{t}}} \quad\text{.......................(11)}.$$

If we now expand $(1-x/t)^{-\frac{1}{2}}$ in powers of x/t, we get

$$\frac{1}{2t}\left[\int_0^t e^{-x}dx + \frac{1}{(2t)}\int_0^t xe^{-x}dx + \ldots + \frac{1.3.5\ldots(2r-1)}{r!\,(2t)^r}\int_0^t x^r e^{-x}dx + \ldots\right]$$

$$= \frac{1}{2t}\left[\gamma(1,t) + \frac{1}{(2t)}\gamma(2,t) + \ldots + \frac{1.3.5\ldots(2r-1)}{r!\,(2t)^r}\gamma(r+1,t) + \ldots\right],$$

where $\gamma(r+1, t)$ is written for the incomplete gamma integral

$$\int_0^t x^r e^{-x}dx.$$

The $(r+1)$th term of the series in square brackets may be written

$$\frac{(r-\frac{1}{2})!}{\sqrt{\pi}\,.\,r!\,t^r}\gamma(r+1,t),$$

or in terms of the *Incomplete Gamma-Function* $I(u,p) = \int_0^{u\sqrt{p+1}} v^p e^{-v}dv \Big/ p!$ we have

$$e^{-t}\int_0^1 e^{tu^2}du = \frac{1}{2t\sqrt{\pi}}\sum_{r=0}^{\infty}\left\{\frac{(r-\frac{1}{2})!}{t^r} I\left(\frac{t}{\sqrt{r+1}}, r\right)\right\}$$

which can be evaluated from the *Tables of the Incomplete* Γ-*function* *.

Since the complete integral $I(\infty, r)$ is equal to unity, we see that the integral in (11) becomes

$$\frac{1}{2t}[1+O(1/t)],$$

which tends to zero as t tends to infinity. This can also be inferred directly from (11), for in the second integral the important part of the integrand when t is large is e^{-x}, and the integral therefore tends to resemble

$$(1-e^{-t})/(2t),$$

which tends to zero as t tends to infinity.

The value of (11) in direct powers of t is the uniformly convergent series

$$e^{-t}\sum_{r=0}^{\infty}\frac{t^r}{r!\,(2r+1)} = e^{-t}F(\tfrac{1}{2}, \tfrac{3}{2}, t) = F(1, \tfrac{3}{2}, -t),$$

* H.M. Stationery Office, 1922.

but as computation would only be feasible from this series with t of the order of unity or lower, it would be better to go back to a direct application of (8), especially when it is remembered that small samples are usually of little value, and values of $(a+b)$ of less than 20 to 25 will be of little practical importance.

Equation (9) is, of course, a direct consequence of the application to the confluent hypergeometric series in (8) of a well-known formula

$$F(\alpha,\gamma,x)=\frac{(\gamma-1)!}{(\gamma-\alpha-1)!}(-x)^{-\alpha}\left\{1-\frac{\alpha(\alpha-\gamma+1)}{x}+\frac{\alpha(\alpha+1)(\alpha-\gamma+1)(\alpha-\gamma+2)}{2!\,x^2}-\dots\right\}$$
$$+\frac{(\gamma-1)!}{(\alpha-1)!}e^{x}x^{\alpha-\gamma}\left\{1+\frac{(1-\alpha)(\gamma-\alpha)}{x}+\frac{(1-\alpha)(2-\alpha)(\gamma-\alpha)(\gamma-\alpha+1)}{2!\,x^2}+\dots\right\},$$

where in our case $\alpha=1$, $\gamma=a+b+1$, $x=-t$. This relation, however, breaks down in our case when $(a+b)$ is a half-integer owing to the second series becoming imaginary. Certain tables of the confluent hypergeometric series have been computed by Airey*, but they are of no use to us in the present investigation, being only calculated for values of our $(a+b+1)$ equal to $\frac{1}{2}$, 1, $1\frac{1}{2}$, 2, 3 and 4, and for positive values of x (i.e. negative values of t).

Second Moment Coefficient of E^2.

Returning now to the form (4) of the distribution, we shall multiply by x^2 and integrate from 0 to 1. We find

$$\mu_2'(E^2)=\int_0^1\frac{e^{-t}}{(b-1)!}x^2(1-x)^{b-1}d\{f^{(b-1)}(x)\}$$
$$=-\int_0^1\frac{e^{-t}}{(b-1)!}\{2x(1-x)^{b-1}-(b-1)x^2(1-x)^{b-2}\}f^{(b-1)}(x)\,dx$$
$$=\int_0^1\frac{e^{-t}}{(b-1)!}\{2(1-x)^{b-1}-4(b-1)x(1-x)^{b-2}$$
$$+(b-1)(b-2)x^2(1-x)^{b-3}\}f^{(b-2)}(x)\,dx$$

on integrating by parts. The part taken between 0 and 1 vanishes at these limits in both cases. Continuing the process, we finally obtain

$$\mu_2'(E^2)=\int_0^1\frac{e^{-t}}{(b-1)!}\{\tfrac{1}{2}(b-1)(b-2)(b-1)!(1-x)^2$$
$$-2(b-1)(b-1)!\,x(1-x)+(b-1)!\,x^2\}f'(x)\,dx$$
$$=\int_0^1 e^{-t}\{\tfrac{1}{2}(b-1)(b-2)-b(b-1)x+\tfrac{1}{2}b(b+1)x^2\}\,d(x^{a+b-1}e^{tx})$$
$$=\tfrac{1}{2}(b-1)(b-2)-b(b-1)e^{-t}\Big[x^{a+b}e^{tx}\Big]_0^1+b(b-1)e^{-t}\int_0^1 x^{a+b-1}e^{tx}dx$$
$$+\tfrac{1}{2}b(b+1)e^{-t}\Big[x^{a+b+1}e^{tx}\Big]_0^1-b(b+1)e^{-t}\int_0^1 x^{a+b}e^{tx}dx$$
$$=1+b(b-1)e^{-t}\int_0^1 x^{a+b-1}e^{tx}dx-b(b+1)e^{-t}\int_0^1 x^{a+b}e^{tx}dx\quad\dots\dots\dots(12).$$

* J. R. Airey: British Association, *Reports* for 1926 and 1927.

If in (12) we put $t = 0$, it becomes

$$\mu_2'(E^2) = 1 + \frac{b(b-1)}{a+b} - \frac{b(b+1)}{a+b+1} = \frac{a(a+1)}{(a+b)(a+b+1)}$$
$$= \frac{n_1(n_1+2)}{(n_1+n_2)(n_1+n_2+2)} = \frac{p^2-1}{N^2-1},$$

agreeing with the direct evaluation of the uncorrected second moment coefficient from the special form (2) of the distribution when $\eta^2 = 0$. On the other hand, when t is infinite the same considerations as were taken into account in the determination of the mean value of E^2 show that the second and third members of (12) vanish, and we have $\mu_2'(E^2) = 1$ when $\eta^2 = 1$.

Evaluating the integrals in (12) by expansion of e^{tx} and integrating term by term, we obtain

$$\mu_2'(E^2) = 1 + \frac{b(b-1)}{a+b} F(1, a+b+1, -t) - \frac{b(b+1)}{a+b+1} F(1, a+b+2, -t) \quad \text{...(13).}$$

Now

$$\sigma^2_{E^2} = \mu_2'(E^2) - (\overline{E^2})^2$$
$$= b(b+1)\left[\frac{1}{a+b} F(1, a+b+1, -t) - \frac{1}{a+b+1} F(1, a+b+2, -t)\right]$$
$$- \frac{b^2}{(a+b)^2}\{F(1, a+b+1, -t)\}^2$$
$$= \frac{b(b+1)}{(a+b)(a+b+1)} F(2, a+b+2, -t) - (1-\overline{E^2})^2 \quad \text{...(14).}$$

This may also, if desired, be expressed as

$$\sigma^2_{E^2} = -\frac{b(b+1)}{a+b}\frac{dF}{dt} - (1-\overline{E^2})^2,$$

where F represents the series $F(1, a+b+1, -t)$, i.e. the same series that occurs in the expression (8) for the mean value of E^2. This series will have been calculated in any case to obtain $\overline{E^2}$, either directly or by means of the forms (9) or (10), and if a table were to be prepared of the function F it should be accompanied, for completeness, by a table of its derivative, which could either be calculated directly from the confluent hypergeometric series in (14) or in terms of the differences of the table of F. Direct calculation by (14) involves computation from the series

$$1 - \frac{2t}{a+b+2} + \frac{3t^2}{(a+b+2)(a+b+3)} - \dots,$$

and will not be feasible if t is large compared with $(a+b)$. If this is the case, the simplest way to get the appropriate form for $\sigma^2_{E^2}$ is to differentiate the series in (9) or (10) with respect to t. We have

$$F = \frac{a+b}{t}\left[1 + \sum_{r=1}^{a+b-2} \frac{(a+b-1)\dots(a+b-r)}{(-t)^r} + \frac{(a+b-1)!}{(-t)^{a+b-1}}(1-e^{-t})\right]$$

$(a+b)$ an integer,

$$= \frac{a+b}{t}\left[1 + \sum_{r=1}^{a+b-\frac{3}{2}} \frac{(a+b-1)\dots(a+b-r)}{(-t)^r} + \frac{(a+b-1)\dots\frac{3}{2}}{(-t)^{a+b-\frac{3}{2}}}\left(1 - e^{-t}\int_0^1 e^{tu^2}du\right)\right]$$

$(a+b)$ a half-integer.

Whence

$$\frac{dF}{dt} = -\frac{a+b}{t^2}\left[1 + \sum_{r=1}^{a+b-1}\frac{(r+1)(a+b-1)\dots(a+b-r)}{(-t)^r} + \frac{(a+b-1)!}{(-t)^{a+b-2}}e^{-t}\left(1+\frac{a+b}{t}\right)\right]$$

$$(a+b) \text{ an integer} \quad \dots\dots(15),$$

$$a+b\left[1 + \sum_{r=1}^{a+b-\frac{3}{2}}\frac{(r+1)(a+b-1)\dots(a+b-r)}{(-t)^r}\right.$$
$$\left. + \frac{(a+b)\dots\frac{3}{2}}{(-t)^{a+b-\frac{3}{2}}}\left\{1 - \left(1+\frac{t}{a+b}\right)e^{-t}\int_0^1 e^{tu^2}du\right\}\right] \quad (a+b) \text{ a half-integer} \quad \dots(16).$$

Case when b is a half-integer.

So far the results for the probability integral, and the mean value and variance of E^2, from the distribution (3) have only been proved for the case when b is an integer. The series $F(-b, a, -tx)$ was then a terminating one, and it was possible to express the distribution in terms of the bth differential coefficient, with regard to x, of the function $f(x) = x^{a+b-1}e^{tx}$. When b is a half-integer, the other possibility, it will be found convenient to use the theory of non-integral differentiation*. The appropriate theory may be briefly outlined as follows: Beginning with the function $f(x)$, let us consider the operation of integrating it repeatedly between the limits 0 and x. The first integral is $\int_0^x f(\xi)\,d\xi$. We have then

$$\left(\int_0^x\right)^2 f(\xi)\,d\xi = \int_0^x dy\int_0^y f(\xi)\,d\xi$$
$$= \int_0^x f(\xi)\,d\xi\int_\xi^x dy$$
$$= \int_0^x (x-\xi)\,f(\xi)\,d\xi.$$

Repeating the operation, we have

$$\left(\int_0^x\right)^3 f(\xi)\,d\xi = \int_0^x dy\int_0^y (y-\xi)\,f(\xi)\,d\xi$$
$$= \int_0^x f(\xi)\,d\xi\int_\xi^x (x-y)\,dy$$
$$= \int_0^x \frac{(x-\xi)^2}{2!}\,f(\xi)\,d\xi.$$

In general
$$\left(\int_0^x\right)^n f(\xi)\,d\xi = \int_0^x \frac{(x-\xi)^{n-1}}{(n-1)!}\,f(\xi)\,d\xi.$$

Putting $n = \frac{1}{2}$, we have

$$\left(\int_0^x\right)^{\frac{1}{2}} f(\xi)\,d\xi = \int_0^x \frac{f(\xi)\,d\xi}{\sqrt{\pi(x-\xi)}}.$$

This expression may be regarded as the definition of the $\frac{1}{2}$th order integral of the function $f(x)$.

* I am indebted to Professor Norbet Wiener for this suggestion and for references to the literature of the subject.

To apply this result let $f(x) = x^{a+b-1}e^{tx}$, as before. Then we may write

$$\frac{d^b}{dx^b} f(x) = \frac{d^{b+\frac{1}{2}}}{dx^{b+\frac{1}{2}}}\left[\int_0^x \frac{\xi^{a+b-1}e^{t\xi}}{\sqrt{\pi(x-\xi)}}\, d\xi\right] = h^{(b+\frac{1}{2})}(x) \quad\text{.............. (17)},$$

where $h(x)$ denotes the function within square brackets.

Returning now to (4), the distribution of x may be written, formally at least, as

$$df = \frac{e^{-t}}{(b-1)!}(1-x)^{b-1} h^{(b+\frac{1}{2})}(x)\, dx \quad\text{.................... (18)},$$

and $b + \frac{1}{2}$ is now an integer.

Nature of the function $h(x)$ and its derivatives.

Putting $v = \xi/x$ in the integral of (17), we have

$$\frac{1}{\sqrt{\pi}}\int_0^x \frac{\xi^{a+b-1}e^{t\xi}}{\sqrt{x-\xi}}\, d\xi = \frac{x^{a+b-\frac{1}{2}}}{\sqrt{\pi}}\int_0^1 v^{a+b-1}(1-v)^{-\frac{1}{2}} e^{txv}\, dv,$$

i.e.

$$h(x) = \frac{(a+b-1)!}{(a+b-\frac{1}{2})!} x^{a+b-\frac{1}{2}} F(a+b,\, a+b+\tfrac{1}{2},\, tx)$$

$$= \frac{(a+b-1)!}{(a+b-\frac{1}{2})!} x^{a+b-\frac{1}{2}} e^{tx} F(\tfrac{1}{2},\, a+b+\tfrac{1}{2},\, -tx) \quad\text{........... (19)},$$

by Kummer's formula, F denoting as before the confluent hypergeometric series. In the first of the two forms given (19) may be differentiated repeatedly without much difficulty, and we have the results

$$h'(x) = \frac{(a+b-1)!}{(a+b-\frac{3}{2})!} x^{a+b-\frac{3}{2}} F(a+b,\, a+b-\tfrac{1}{2},\, tx) \quad\text{.................... (20)},$$

and in general

$$h^{(r)}(x) = \frac{(a+b-1)!}{(a+b-\frac{1}{2}-r)!} x^{a+b-\frac{1}{2}-r} F(a+b,\, a+b+\tfrac{1}{2}-r,\, tx)$$

$$= \frac{(a+b-1)!}{(a+b-\frac{1}{2}-r)!} x^{a+b-\frac{1}{2}-r} e^{tx} F(\tfrac{1}{2}-r,\, a+b+\tfrac{1}{2}-r,\, -tx) \quad\text{...... (21)}.$$

Putting $r = b + \frac{1}{2}$ and substituting in (18), we get immediately the form (3) of the distribution. An alternative form for $h(x)$, useful for computation purposes when t is large, may be obtained by putting $x - \xi = u^2/2t$ in the integral leading to (19). We then have

$$h(x) = \sqrt{\frac{2}{\pi t}}\, x^{a+b-1} e^{tx} \int_0^{\sqrt{2tx}} \left(1 - \frac{u^2}{2tx}\right)^{a+b-1} e^{-\frac{1}{2}u^2}\, du.$$

On expanding $(1 - u^2/2tx)^{a+b-1}$ by the binomial theorem, and integrating term by term, we have the result

$$h(x) = \frac{2}{\sqrt{t}} x^{a+b-1} e^{tx}\left[m_0(\sqrt{2tx}) - \frac{a+b-1}{2tx} m_2(\sqrt{2tx}) + \frac{1.3.(a+b-1)(a+b-2)}{2!(2tx)^2} m_4(\sqrt{2tx}) - \ldots\right] \quad\text{...... (22)},$$

where $m_{2r}(\sqrt{2tx})$ denotes the $(2r)$th Incomplete Normal Moment Function

$$m_{2r}(\sqrt{2tx}) = \frac{1}{\sqrt{2\pi}}\int_0^{\sqrt{2tx}} u^{2r} e^{-\frac{1}{2}u^2} du \Big/ (2r-1)(2r-3)\dots 1,$$

$$m_0(\sqrt{2tx}) = \frac{1}{\sqrt{2\pi}}\int_0^{\sqrt{2tx}} e^{-\frac{1}{2}u^2} du.$$

m_0 may therefore be obtained from Sheppard's tables*, and the even m's required are tabulated up to m_{12}*. The upper limit of $m_{2r}(\sqrt{2tx})$, when t becomes infinite, is in all cases 0·5.

Probability Integral.

Returning now to (18) the probability integral is

$$I = \frac{e^{-t}}{(b-1)!}\int_0^x (1-x)^{b-1} h^{(b+\frac{1}{2})}(x)\, dx.$$

This may be integrated successively by parts, putting $h^{(b+\frac{1}{2})}(x)\,dx = d\{h^{(b-\frac{1}{2})}(x)\}$, etc., and we find

$$\begin{aligned} I &= e^{-t}\sum_{r=1}^{b-\frac{1}{2}} \frac{(1-x)^{r-\frac{1}{2}}}{(r-\frac{1}{2})!} h^{(r)}(x) + \frac{e^{-t}}{\sqrt{\pi}}\int_0^x \frac{h'(x)\,dx}{\sqrt{1-x}} \\ &= \frac{(a+b-1)!}{e^{t(1-x)}}\sum_{r=1}^{b-\frac{1}{2}}\left\{\frac{x^{a+b-\frac{1}{2}-r}(1-x)^{r-\frac{1}{2}}}{(a+b-\frac{1}{2}-r)!\,(r-\frac{1}{2})!} F(\tfrac{1}{2}-r,\, a+b+\tfrac{1}{2}-r,\, -tx)\right\} \\ &\qquad + \frac{e^{-t}}{\sqrt{\pi}}\int_0^x \frac{h'(x)\,dx}{\sqrt{1-x}} \quad \text{......(23)}, \end{aligned}$$

on substituting for $h^{(r)}(x)$ from (21). This is the analogous form to our result (5), and may be completed by evaluating the integral in (23).

$h'(x)$ may be written, from (20),

$$h'(x) = \sum_{r=0}^{\infty}\left\{\frac{(a+b+r-1)!}{(a+b+r-\frac{3}{2})!}\frac{t^r}{r!} x^{a+b+r-\frac{3}{2}}\right\}.$$

Consequently

$$\begin{aligned} \frac{e^{-t}}{\sqrt{\pi}}\int_0^x \frac{h'(x)\,dx}{\sqrt{1-x}} &= \frac{e^{-t}}{\sqrt{\pi}}\sum_{r=0}^{\infty}\left\{\frac{(a+b+r-1)!}{(a+b+r-\frac{3}{2})!}\frac{t^r}{r!}\int_0^x x^{a+b+r-\frac{3}{2}}(1-x)^{-\frac{1}{2}}\,dx\right\} \\ &= e^{-t}\sum_{r=0}^{\infty}\left\{\frac{t^r}{r!} I_x(a+b+r-\tfrac{1}{2},\, \tfrac{1}{2})\right\}, \end{aligned}$$

where $I_x(p, q)$ is the incomplete Beta function

$$\frac{(p+q-1)!}{(p-1)!\,(q-1)!}\int_0^x x^{p-1}(1-x)^{q-1}dx.$$

The required integral can be expressed, therefore, in a simple series of incomplete Beta functions, which can be obtained for small values of $(a+b)$ from the tables prepared in the Galton Laboratory, and now at press†. For the special case of

* *Tables for Statisticians*, Part I, pp. 2 and 23; Part II, p. 147.

† See also *Biometrika*, Vol. XXII. pp. 274—283, and *Tables for Statisticians*, Part II, 1931, Table XXV.

p or q equal to $\frac{1}{2}$, the incomplete Beta function can be expressed in terms of the symmetrical integral as follows:

$$I_x(p, \tfrac{1}{2}) = 1 - 2I_\theta(2p-1),$$

where
$$I_\theta(2p-1) = \int_0^\theta \cos^{2p-1}\theta\, d\theta \Big/ \left\{2\int_0^{\frac{\pi}{2}} \cos^{2p-1}\theta\, d\theta\right\},$$

and $\theta = \cos^{-1}\sqrt{x}$. The best method of evaluating $I_\theta(2p-1)$, in the absence of tables, is one I gave some years ago* in terms of a series of powers of $1/(2p-1)$.

When $x = 1$, $I_1(pq) = 1$, and therefore $\frac{e^{-t}}{\sqrt{\pi}}\int_0^1 \frac{h'(x)\,dx}{\sqrt{1-x}} = e^{-t}\sum_{r=0}^{\infty}\left(\frac{t^r}{r!}\right) = 1.$

Mean value of E^2.

Taking the form (18) of the distribution, let us multiply by x and integrate from 0 to 1. The procedure is exactly similar to that in the earlier part of the paper, where b was considered an integer, and the part taken between the limits 0 and 1 on integrating by parts vanishes in every case. Finally we are left with

$$\begin{aligned}
E^2 &= \frac{e^{-t}}{(b-1)!}\int_0^1 \{-(b-\tfrac{1}{2})(b-1)(b-2)\ldots\tfrac{3}{2}(1-x)^{b-(b-\frac{1}{2})} \\
&\qquad\qquad + (b-1)(b-2)\ldots\tfrac{1}{2}x(1-x)^{b-(b+\frac{1}{2})}\}\, h'(x)\,dx \\
&= e^{-t}\int_0^1 \left\{\frac{-(b-\frac{1}{2})(1-x)^{\frac{1}{2}}}{\frac{1}{2}\sqrt{\pi}} + \frac{x}{\sqrt{\pi}(1-x)}\right\} h'(x)\,dx \\
&= \frac{e^{-t}}{\sqrt{\pi}}\left\{\int_0^1 \frac{-(2b-1)h'(x)\,dx}{\sqrt{1-x}} + \int_0^1 \frac{2bx\,h'(x)\,dx}{\sqrt{1-x}}\right\} \\
&= \frac{e^{-t}}{\sqrt{\pi}}\int_0^1 \frac{h'(x)\,dx}{\sqrt{1-x}} - \frac{2b\,e^{-t}}{\sqrt{\pi}}\int_0^1 \sqrt{1-x}\,h'(x)\,dx \\
&= 1 - \frac{2b\,e^{-t}}{\sqrt{\pi}}\int_0^1 \sqrt{1-x}\,h'(x)\,dx \\
&= 1 - \frac{2b\,e^{-t}}{\sqrt{\pi}}\sum_{r=0}^{\infty}\left\{\frac{(a+b+r-1)!\,t^r}{(a+b+r-\frac{3}{2})!\,r!}\int_0^1 x^{a+b+r-\frac{3}{2}}(1-x)^{\frac{1}{2}}\,dx\right\} \quad \text{from (20)} \\
&= 1 - b\,e^{-t}\sum_{r=0}^{\infty}\frac{t^r}{r!(a+b+r)} \\
&= 1 - \frac{b\,e^{-t}}{a+b}F(a+b, a+b+1, t).
\end{aligned}$$

Finally
$$\overline{E^2} = 1 - \frac{b}{a+b}F(1, a+b+1, -t).$$

Thus our result (8) for the mean value of E^2 is shown to hold in general for all values, integral and half-integral, of b. The alternative forms (9) and (10), for large t, follow as a matter of course.

* *Biometrika*, Vol. XVII. 1925, pp. 469—472. See *Tables for Statisticians*, Part II, 1931, pp. ccxxi and 289. Note that this formula is applicable generally for any value of N (our $2p-1$) and should be regarded as superseding the former method and Table XLV.

Second Moment Coefficient of E^2.

We repeat the former procedure of multiplying by x^2 and integrating successively by parts, stopping when we reach an integral involving $h'(x)$. The parts between the limits vanish, and we have

$$\begin{aligned}\mu_2'(E^2) = \frac{e^{-t}}{(b-1)!}\int_0^1 \{&(b-\tfrac{1}{2})(b-\tfrac{3}{2})(b-1)(b-2)\ldots\tfrac{5}{2}(1-x)^{b-(b-\frac{3}{2})}\\ &-2(b-\tfrac{1}{2})(b-1)(b-2)\ldots\tfrac{3}{2}x(1-x)^{b-(b-\frac{1}{2})}\\ &+(b-1)(b-2)\ldots\tfrac{1}{2}x^2(1-x)^{b-(b+\frac{1}{2})}\}\,h'(x)\,dx\\ = e^{-t}\int_0^1 &\left\{\frac{(b-\frac{1}{2})(b-\frac{3}{2})}{\frac{1}{2}\cdot\frac{3}{2}\cdot\sqrt{\pi}}(1-x)^{\frac{3}{2}} - \frac{2(b-\frac{1}{2})}{\frac{1}{2}\sqrt{\pi}}x(1-x)^{\frac{1}{2}} + \frac{x^2(1-x)^{-\frac{1}{2}}}{\sqrt{\pi}}\right\}h'(x)\,dx\\ = \frac{4e^{-t}}{3\sqrt{\pi}}\int_0^1 &\{(b-\tfrac{1}{2})(b-\tfrac{3}{2}) - 2b(b-\tfrac{1}{2})x + b(b+1)x^2\}\frac{h'(x)\,dx}{\sqrt{1-x}}\\ = \frac{e^{-t}}{\sqrt{\pi}}&\left[\int_0^1\frac{h'(x)\,dx}{\sqrt{1-x}} - 4b\int_0^1\sqrt{1-x}\,h'(x)\,dx + \tfrac{4}{3}b(b+1)\int_0^1(1-x)^{\frac{3}{2}}h'(x)\,dx\right].\end{aligned}$$

The first and second of these integrals have already been evaluated, while the third may be obtained directly from (20) by a suitable modification of the procedure already outlined. We then have

$$\begin{aligned}\mu_2'(E^2) &= 1 - 2be^{-t}\sum_{r=0}^{\infty}\frac{t^r}{r!\,(a+b+r)} + b(b+r)e^{-t}\sum_{r=0}^{\infty}\frac{t^r}{(a+b+r)(a+b+r+1)}\\ &= 1 - \frac{2b}{a+b}e^{-t}F(a+b,\,a+b+1,\,t) + \frac{b(b+1)}{(a+b)(a+b+1)}e^{-t}F(a+b,\,a+b+2,\,t)\\ &= 1 - \frac{2b}{a+b}F(1,\,a+b+1,\,-t) + \frac{b(b+1)}{(a+b)(a+b+1)}F(2,\,a+b+2,\,-t).\end{aligned}$$

This, though not of precisely the same form as (13), is seen to be equivalent to it when we derive $\sigma^2_{E^2}$, for subtracting the square of the mean

$$(\overline{E^2})^2 = 1 - \frac{2b}{a+b}F(1,\,a+b+1,\,-t) + \frac{b^2}{(a+b)^2}F^2(1,\,a+b+1,\,-t),$$

we see that

$$\sigma^2_{E^2} = \frac{b(b+1)}{(a+b)(a+b+1)}F(2,\,a+b+2,\,-t) - (1-\overline{E^2})^2,$$

which agrees exactly with (14). This result, therefore, holds generally for the variance of E^2 whether b is an integer or a half-integer.

Limiting values of Mean and Second Moment of E^2.

In conclusion it will be of interest to show the identity of the limiting value of the expressions (8) and (14) with the corresponding moments of Fisher's distribution (C)*. To establish this we require to find the limits to which the moments of bE^2 tend as b is made indefinitely large.

Now $$\overline{E^2} = 1 - \left(1+\frac{a}{b}\right)^{-1}\left\{1 - \frac{t}{b}\left(1+\frac{a+1}{b}\right)^{-1} + \frac{t^2}{b^2} + O\left(\frac{1}{b^3}\right)\right\} \text{ from (8)}$$

$$= \frac{a+t}{b} - \frac{(a+t)^2+t}{b^2} + O\left(\frac{1}{b^3}\right) \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots(24).$$

* *Proc. Roy. Soc.* A, 121, 1928, p. 663.

Hence $$\operatorname*{Lt}_{b\to\infty} \overline{bE^2} = a + t \quad\text{.......(25).}$$

Also

$$\sigma^2_{E^2} = \left(1+\frac{1}{b}\right)\left(1+\frac{a}{b}\right)^{-1}\left(1+\frac{a+1}{b}\right)^{-1}\left\{1-\frac{2t}{b}\left(1+\frac{a+2}{b}\right)^2+\frac{3t^2}{b^2}+O\left(\frac{1}{b^3}\right)\right\}$$
$$-(1-\overline{E^2})^2 \text{ from (14)}$$
$$=1-\frac{2(a+t)}{b}+\frac{3(a+t)^2+a+4t}{b^2}+O\left(\frac{1}{b^3}\right)-(1-\overline{E^2})^2.$$

But $$(1-\overline{E^2})^2 = 1-\frac{2(a+t)}{b}+\frac{3(a+t)^2+2t}{b^2}+O\left(\frac{1}{b^3}\right) \text{ from (24).}$$

Therefore $$\sigma^2_{E^2} = \frac{a+2t}{b^2}+O\left(\frac{1}{b^3}\right),$$

and $$\operatorname*{Lt}_{b\to\infty} \sigma^2_{bE^2} = a+2t \quad\text{.......(26).}$$

Now consider the (C) distribution of Fisher. Writing $\frac{1}{2}B^2 = x$, $\frac{1}{2}\beta^2 = t$ and $\frac{1}{2}n_1 = a$, it takes the form

$$df = \frac{x^{a-1}}{(a-1)!}e^{-x-t}\left\{1+\frac{(tx)}{a}+\frac{(tx)^2}{2!\,a(a+1)}+\ldots\right\}dx \quad\text{.........(27),}$$

which we may put in the form

$$df = \left(\frac{x}{t}\right)^{\frac{a-1}{2}} e^{-x-t} I_{a-1}(2\sqrt{xt})\,dx,$$

where I denotes the Bessel Function of imaginary argument.

The range of x in (27) is from 0 to ∞, and the integral over this range is unity.

The *moment generating function* M is therefore defined by

$$M = \int_0^\infty e^{hx}\,df,$$

since the coefficient of $h^r/r!$ in the expansion of this expression is $\mu_r'(x)$, the rth moment coefficient about the origin.

Thus $$M = \int_0^\infty \frac{x^{a-1}}{(a-1)!}e^{-t}e^{-x(1-h)}\left\{1+\frac{(tx)}{a}+\frac{(tx)^2}{2!\,a(a+1)}+\ldots\right\}dx,$$

or $$M = \int_0^\infty \left(\frac{x}{t}\right)^{\frac{a-1}{2}} e^{-t}e^{-x(1-h)} I_{a-1}(2\sqrt{xt})\,dx.$$

Let us now change the variable, writing $x(1-h) = x'$ and at the same time writing t' for $t/(1-h)$ in order to keep the series part in its present form. We then have

$$M = \frac{e^{th/1-h}}{(1-h)^a}\int_0^\infty \frac{x'^{a-1}}{(a-1)!}e^{-x'-t'}\left\{1+\frac{(t'x')}{a}+\frac{(t'x')^2}{2!\,a(a+1)}+\ldots\right\}dx'$$
$$=\frac{e^{th/1-h}}{(1-h)^a}\int_0^\infty \left(\frac{x'}{t'}\right)^{\frac{a-1}{2}} e^{-x'-t'} I_{a-1}(2\sqrt{x't'})\,dx'$$
$$=\frac{e^{th/1-h}}{(1-h)^a} \quad\text{.......(28),}$$

since the integral is now of the same form as (27).

Further, if we write

$$K = \log M = \kappa_1 h + \kappa_2 \frac{h^2}{2!} + \kappa_3 \frac{h^3}{3!} + \dots,$$

on expanding in powers of h, then κ_r is the rth *semi-invariant* of the distribution of x. κ_1 is equal to μ_1', while

$$\kappa_2 = \mu_2, \quad \kappa_3 = \mu_3,$$

$$\kappa_4 = \mu_4 - 3\mu_2^2 = \mu_2^2(\beta_2 - 3),$$

and so on.

We have therefore only to take the logarithm of (28) and find the term in $h^r/r!$ in its expansion.

Now $$K = \frac{th}{1-h} - a\log(1-h),$$

and the term in $h^r/r!$ in this is simply

$$\kappa_r(x) = (rt + a)(r-1)! \quad \text{.........................(29).}$$

This is the general rth semi-invariant of the distribution (27), and is a generalisation of the Type III result $a(r-1)!$ which holds when t is zero, as is otherwise evident from inspection of (27)*.

In particular the mean value of x, i.e. of $\underset{b\to\infty}{\mathrm{Lt}}\,(bE^2)$, is obtained from $\kappa_r(x)$ by putting $r=1$, and is $a+t$ (see (25)), while the second moment coefficient, σ^2 (identical with $\kappa_2(x)$), is equal to $a+2t$ (see (26)). This establishes the relations sought.

We have noted already that Fisher's (A) distribution tends, like (C), to the common limit (B) as n_2 increases without limit. We are not concerned here with the properties of the (A) distribution, but it is perhaps relevant to show how the results of the previous paper† on the mean and variance of the (A) distribution also tend, in the limit, to the values (25) and (26) just deduced. In terms of the correlation ratio we have

$$\overline{E^2} = \frac{a}{a+b} + \frac{b\eta^2}{a+b+1} F(1, 1, a+b+2, \eta^2)$$

from equation (14) of the former paper.

Now write $b\eta^2 = \frac{1}{2}\beta^2 = t$ in our notation, and we have

$$\overline{E^2} = \frac{a}{b}\left(1+\frac{a}{b}\right)^{-1} + \frac{t}{b}\left(1+\frac{a+1}{b}\right)^{-1}\left\{1+\frac{t}{b^2}+O\left(\frac{1}{b^3}\right)\right\}$$

$$= \frac{a+t}{b} - \frac{a^2+(a+1)t}{b^2} + O\left(\frac{1}{b^3}\right) \quad \text{..............................(30).}$$

Hence $$\underset{b\to\infty}{\mathrm{Lt}}\,b\overline{E^2} = a+t.$$

* We have here another case of a Bessel Function distribution, the law for the semi-invariants of which is even simpler than that of McKay, given in *Biometrika*, Vol. XXIV. 1932, pp. 39—44.

† J. Wishart: *Biometrika*, Vol. XXII. 1931, pp. 353—361.

Further, from equation (20) of the former paper we have

$$\begin{aligned}\sigma^2_{E^2} &= \frac{b(b+1)(1-\eta^2)^2}{(a+b)(a+b+1)} F(2, 2, a+b+2, \eta^2) - (1-\overline{E^2})^2 \\ &= \left(1+\frac{1}{b}\right)\left(1-\frac{t}{b}\right)^2\left(1+\frac{a}{b}\right)^{-1}\left(1+\frac{a+1}{b}\right)^{-1}\left\{1+\frac{4t}{b^2}+O\left(\frac{1}{b^3}\right)\right\} - (1-\overline{E^2})^2 \\ &= 1 - \frac{2(a+t)}{b} + \frac{3a^2+a+4at+4t+t^2}{b^2} + O\left(\frac{1}{b^3}\right) - (1-\overline{E^2})^2.\end{aligned}$$

$(1-\overline{E^2})^2$ is obtained from (30), and we have

$$\sigma^2_{E^2} = \frac{a+2t}{b^2} + O\left(\frac{1}{b^3}\right),$$

and

$$\underset{b\to\infty}{\mathrm{Lt}}\ \sigma^2_{bE^2} = a + 2t.$$

In both cases the results are identical with those already deduced from the (C) distribution.

Summary.

Beginning with a statement of the nature of the distributions, for three distinct cases, of the square of the multiple correlation ratio in samples from a normal population, the paper goes on to consider in detail the third of these, namely that appropriate to the case where the array totals are supposed the same for all arrays. Expressions are reached for the probability integral of the distribution, and for the mean value and variance of the square of the sample correlation ratio. It is shown finally that the mean and variance tend, in common with the analogous results for the other general distribution previously studied, to the corresponding parameters of Fisher's limiting distribution (B), as the size of the sample is increased without limit, and the general semi-invariant of this limiting distribution is given.

ON THE PROBABILITY THAT TWO INDEPENDENT DISTRIBUTIONS OF FREQUENCY ARE REALLY SAMPLES FROM THE SAME PARENT POPULATION.

By KARL PEARSON.

1. LET there be v categories in either sample and suppose that the parent population has the same v categories, and that the chance of drawing an individual from the sth category of the parent population is p_s, where s may be 1, 2, 3 ... v. Let the category contents of the two samples of sizes N and N' be respectively

$$n_1, n_2, n_3, \ldots n_s, \ldots n_v,$$

and

$$n_1', n_2', n_3', \ldots n_s', \ldots n_v'.$$

Then I proved in a paper published in 1911*, that if

$$\chi^2 = \overset{s=v}{\underset{s=1}{S}} \frac{NN'}{N+N'} \left(\frac{n_s}{N} - \frac{n_s'}{N'}\right)^2 \frac{1}{p_s} \qquad \text{(i)},$$

the frequency distribution of χ^2 would be given by

$$y = y_0 e^{-\frac{1}{2}\chi^2} (\tfrac{1}{2}\chi^2)^{\frac{1}{2}(v-3)} \qquad [d(\tfrac{1}{2}\chi^2)] \qquad \text{(ii)},$$

and P, the probability of χ^2 not falling short of a given value, would be found by entering the (χ^2, P) tables under that value of χ^2 and with $n' = v$.

If the parent population be known, or we are testing whether the two samples are likely to have been drawn from a hypothetical population, the problem is perfectly straightforward, because the series p_s will be given; and the answer may be found by fairly easy arithmetical work. When, however, the parent population is unknown, and the question put to us is—Are the two samples likely to have been drawn from some unspecified parent population?—the answer is not so easy to provide.

Unfortunately the answer I gave in 1911 was not the correct one. I wrote:

"Now the best hypothesis as to the constitution of this [the parent] population, on the assumption that both frequencies are random samples of it, will be that its sth frequency class is that indicated by the combined two samples†."

In other words, I suggest that p_s should be taken equal to $(n_s + n_s')/(N + N')$. This was not a bad suggestion‡, if the samples were considerable in size, but it is in no way the "best" hypothesis. If we adopt it, then

$$\chi^2 = \frac{1}{NN'} \overset{s=v}{\underset{s=1}{S}} \frac{(N'n_s - Nn_s')^2}{n_s + n_s'} \qquad \text{(iii)},$$

* *Biometrika*, Vol. VIII, pp. 250—254. † *loc. cit.* p. 252.

‡ This use of the sample values for the unknown parent population values is so usual in the theory of errors that it frequently escapes comment; it is really only legitimate in the cases of large samples.

and the table of the two samples can be arranged as a biserial contingency table, and it has unfortunately come to be spoken of as such. The true form (i) cannot be represented as a contingency table, and (iii) will, as a rule, not give a contingency table for any other pair of samples which help to make up the complement of χ^2's involved in P. It has led to many students forgetting what the $n_s + n_s'$ stands for, i.e. an approximation more or less adequate for the unknown $(N + N')p_s$.

If we try to think over what the words "best hypothesis" mean in this matter, ought we not to interpret them as signifying that hypothesis as to the p_s's which will give the highest probability of the two samples being draw from the same population? Surely, if we are asking whether the two samples are likely to have been drawn from some unknown parent population, we ought to choose for that unknown population the one that makes the probability P of their common sampledom a maximum, or the value of χ^2 as small as possible.

Now it is quite possible to determine this system of p_s's. Of course when found they may contradict some other experience we have had. But here in this problem we are supposed to start with no past experience, i.e. with quite unknown p_s's. Had we some previous experience of the latter we should have to discover not "the most likely" but what I have termed "the most reasonable" values of the p_s series*.

Proceeding to the determination of the most likely values of the p_s series, we have to make

$$\chi^2 = \nu \mathop{S}_{s=1}^{s=v} \left(\frac{n_s}{N} - \frac{n_s'}{N'}\right)^2 \frac{1}{p_s},$$

where $\nu = NN'/(N + N')$ a minimum, subject to the conditions that:

(*a*)
$$\mathop{S}_{s=1}^{s=v} (p_s) = 1,$$

and

(*b*) the values obtained for the p_s's are possible as probabilities, i.e. they must all be positive and less than unity.

Following the usual method with an indeterminate multiplier λ:

$$0 = -\nu \mathop{S}_{s=1}^{s=v} \left(\frac{n_s}{N} - \frac{n_s'}{N'}\right)^2 \frac{1}{p_s^2}\,\delta p_s,$$

$$0 = \mathop{S}_{s=1}^{v} (\delta p_s),$$

or we obtain the series of equations

$$\nu \left(\frac{n_s}{N} - \frac{n_s'}{N'}\right)^2 \Big/ p_s^2 + \lambda = 0.$$

Multiply by p_s and sum, we find

$$\chi^2_{\text{min.}} + \lambda = 0.$$

* On the determination of the "most likely" and "most reasonable" values of the constants of a parent population see *Tables for Statisticians and Biometricians*, Part II, pp. clxxi *et seq.*

Hence

$$p_s^2 = \frac{\nu}{\chi^2_{\text{min.}}}\left(\frac{n_s}{N} - \frac{n_s'}{N'}\right)^2.$$

Taking the square root of this we have a plus and minus root, and by the nature of p_s we must take the positive result. In other words

$$p_s = \sqrt{\frac{\nu}{\chi^2_{\text{min.}}}}\left(\frac{n_s}{N} \sim \frac{n_s'}{N'}\right),$$

the quantity in brackets being given the positive sign.

Sum the result just obtained and we have

$$1 = \sqrt{\frac{\nu}{\chi^2_{\text{min.}}}} \overset{s=v}{\underset{s=1}{S}}\left(\frac{n_s}{N} \sim \frac{n_s'}{N'}\right).$$

Thus

$$\chi^2_{\text{min.}} = \nu\left\{\overset{s=v}{\underset{s=1}{S}}\left(\frac{n_s}{N} \sim \frac{n_s'}{N'}\right)\right\}^2 \quad \text{........................(iv)},$$

and

$$p_s = \frac{\left(\frac{n_s}{N} \sim \frac{n_s'}{N'}\right)}{\overset{s=v}{\underset{s=1}{S}}\left(\frac{n_s}{N} \sim \frac{n_s'}{N'}\right)} \quad \text{.(v)}.$$

These values satisfy all the requirements of the problem. p_s is always positive and less than unity, and $\overset{s=v}{\underset{s=1}{S}}(p_s) = 1$. Further, the χ^2 obtained is a minimum and not a maximum because we can choose the p_s series so that χ^2 can be as large as we please.

Accordingly we have chosen a parent population which gives us the best chance that the two samples are drawn from the same parent population.

We will now illustrate this numerically.

2. The following data have been extracted by R. A. Fisher* from Tocher's Scottish returns for the children of a certain locality:

TABLE I.

Hair Colour.

	Fair	Red	Medium	Dark	Jet Black	Totals
Boys	592	119	849	504	36	2100
Girls	544	97	677	451	14	1783
Totals	1136	216	1526	955	50	3883

* *Statistical Methods for Research Workers*, 3rd ed. p. 85.

Here $N = 2100$, $N' = 1783$ and the proportional frequencies are:

	Fair	Red	Medium	Dark	Jet Black	Totals
Boys $\left(\frac{n_s}{N}\right)$	·281,9048	·056,6667	·404,2857	·240,0000	·017,1428	1·000,0000
Girls $\left(\frac{n_s'}{N'}\right)$	·305,1038	·054,4027	·379,6971	·252,9445	·007,8519	1·000,0000
$\frac{n_s}{N} - \frac{n_s'}{N'}$	− ·023,1990	·002,2640	·024,5886	− ·012,9445	·009,2909	—

Further: $$NN'/(N+N') = \nu = 964{\cdot}2802.$$

Everything is now ready for substitution in (i), when we have selected our p_s series. Fisher takes for his p_s series the values $(n_s + n_s')/(N+N')$, which are obtained by dividing the third row of figures by its total 3883. Let these be termed p_s''s. We have

$$p_1' = {\cdot}292{,}5573, \quad p_2' = {\cdot}055{,}6271, \quad p_3' = {\cdot}392{,}9951, \quad p_4' = {\cdot}245{,}9439, \quad p_5' = {\cdot}012{,}8766 \quad \ldots\ldots\text{(vi)}.$$

If now we square the last line of the table above, divide each square by its appropriate p_s' from (vi), add and multiply the result by ν, we find

$$\chi^2 = 10{\cdot}4674.$$

Fisher gives practically an equivalent $\chi^2 = 10{\cdot}468$. By our method we have five categories—no question arises of degrees of freedom—and we look out the (χ^2, P) table under $\chi^2 = 10{\cdot}4674$ and $n' = 5$ and find $P = {\cdot}034$. This agrees with Fisher who says the value of P lies between ·02 and ·05. He concludes "that the sex difference in the classification by hair colours is probably significant as judged by this district alone."

We ask, however, whether another parent population could not be found contradicting this result.

Turning back to the table at the top of the page we add *without regard to sign* its third row of figures and find for its total

$$\mathop{S}_{s=1}^{s=5}\left(\frac{n_s}{N} \sim \frac{n_s'}{N'}\right) = {\cdot}072{,}2870,$$

whence by (ν) we have

$$p_1 = {\cdot}320{,}9291, \quad p_2 = {\cdot}031{,}3196, \quad p_3 = {\cdot}340{,}1524, \quad p_4 = {\cdot}179{,}0709, \quad p_5 = {\cdot}128{,}5280 \quad \ldots\ldots\text{(vii)}.$$

Now here are another series of p's, which will not like the p_s''s in (vi) lie between the values found for the two samples. But have we any reason to suppose the parent population must have relative frequencies lying between those of the two samples? All we can say, if we are in complete ignorance of the parent population, and are determining "the most likely" values of its proportional frequencies, that we ought to use (vii) rather than (vi). We can either use (i) as we have already squared the items of the third row of the table above, or more briefly use (iv), we have

$$\chi^2_{\text{min.}} = 964{\cdot}2802\,({\cdot}072{,}2870)^2 = 5{\cdot}0388,$$

a value less than half that given when we use the combined samples' value

$$(n_s + n_s')/(N + N')$$

to determine p_s. This value of χ^2 leads to $P = \cdot 284$, from which we should conclude that it would be quite reasonable to suppose *no* sexual difference in regard to hair colour of boys and girls in this particular locality.

This illustration is not given to prove that there is no sexual difference in hair colour, but as a warning against placing too great a trust in the χ^2 method of estimating the difference of two samples by the spurious contingency method, i.e. the method which without thinking of the parent population, and the nature of the assumption $p_s = (n_s + n_s')/(N + N')$ uses (iii) as an invariably safe test. There are clearly numerous parent populations between (vi) and (vii) which would admit of the two samples being reasonably considered to have arisen from the same population.

3. We can extend the conception of the first section of this paper to a more general problem. Suppose we have a population involving two characteristics A and B classified into u and v categories respectively, and let the chance of an individual being drawn from the sth category of A be p_s and from the tth category of B be q_t. Further, let the chance of an individual being drawn combining both characteristics be α_{st}, where α_{st} will not be equal to $p_s \times q_t$ unless the characteristics are independent. Now we can represent this population in the form of the table

	B_1	B_2	...	B_t	...	B_v	
A_1	α_{11},	α_{12},	...	α_{1t},	...	α_{1v},	p_1
A_2	α_{21},	α_{22},	...	α_{2t},	...	α_{2v},	p_2
...	...	...	...	...	...	...	...
A_s	α_{s1},	α_{s2},	...	α_{st},	...	α_{sv},	p_s
...	...	...	...	...	...	...	...
A_u	α_{u1},	α_{u2},	...	α_{ut},	...	α_{uv},	p_u
	q_1	q_2	...	q_t	...	q_v	1

Now suppose a sample of size N be drawn from the above population as parent population, and let the distribution in the $u \times v$ cells be represented by the scheme below:

n_{11}	n_{12}	...	n_{1t}	...	n_{1v}	$n_{1.}$
n_{21}	n_{22}	...	n_{2t}	...	n_{2v}	$n_{2.}$
...	...	...	...	...	...	...
n_{s1}	n_{s2}	...	n_{st}	...	n_{sv}	$n_{s.}$
...	...	...	...	...	...	...
n_{u1}	n_{u2}	...	n_{ut}	...	n_{uv}	$n_{u.}$
$n_{.1}$	$n_{.2}$	...	$n_{.t}$	...	$n_{.v}$	N

Then the mean square contingency of such a sample is defined as

$$\phi^2 = \frac{1}{N} \mathop{S}_{s=1}^{s=u} \mathop{S}_{t=1}^{t=v} \frac{(n_{st} - N\alpha_{st})^2}{N\alpha_{st}} = \frac{1}{N} \mathop{S}_{s=1}^{s=u} \mathop{S}_{t=1}^{t=v} \left(\frac{n_{st}^2}{N\alpha_{st}^2}\right) - 1.$$

Similarly, if χ^2 be taken as $N\phi^2$, we have

$$\chi^2 + N = \mathop{S}_{s=1}^{s=u} \mathop{S}_{t=1}^{t=v} \left(\frac{n_{st}^2}{N\alpha_{st}}\right) \quad \text{.............................(viii).}$$

Now if there be no correlation between the variates in the parent population, i.e. $\alpha_{st} = p_s \times q_t$, then on certain assumptions with regard to the size of n_{st}, χ^2 and therefore ϕ^2/N *as thus* defined will be distributed according to the law

$$y = y_0 e^{-\frac{1}{2}\chi^2} (\tfrac{1}{2}\chi^2)^{\frac{1}{2}(uv-3)} \quad \text{.............................(ix),}$$

for the n_{st}'s being independent, it does not matter whether we arrange them in the form of the above table, or in a single row or column.

Now, suppose we have no knowledge whatever of the parent population, then what are the best values to give the unknown α_{st}'s?

Clearly the only relation binding on their choice is $\mathop{S}_{s=1}^{s=u} \mathop{S}_{t=1}^{t=v} (\alpha_{st}) = 1$.

Let us find the minimum value of χ^2 subject to this condition; we have dropping the double summation sign for brevity

$$0 = -S\left(\frac{n_{st}^2 \delta\alpha_{st}}{N\alpha_{st}^2}\right),$$

$$0 = S(\delta\alpha_{st}).$$

Using an indeterminate multiplier λ, we have

$$-\frac{n_{st}^2}{N\alpha_{st}^2} + \lambda = 0,$$

for all values of s and t.

Multiplying by α_{st} and summing

$$S\left(\frac{n_{st}^2}{N\alpha_{st}}\right) = \lambda, \quad \text{or} \quad \lambda = \chi^2_{\text{min.}} + N.$$

Further:
$$\alpha_{st} = \frac{n_{st}}{\sqrt{N\lambda}} = \frac{n_{st}}{\sqrt{N}\sqrt{\chi^2_{\text{min.}} + N}},$$

and accordingly:

$$S\left(\frac{n_{st}^2}{Nn_{st}}\right)\sqrt{N}\sqrt{\chi^2_{\text{min.}} + N} = \chi^2_{\text{min.}} + N,$$

$$S\left(\frac{n_{st}}{\sqrt{N}}\right) = \sqrt{N} = \sqrt{\chi^2_{\text{min.}} + N},$$

or
$$\chi^2_{\text{min.}} = 0, \quad \text{and} \quad \alpha_{st} = \frac{n_{st}}{N},$$

from which

$$p_s = \mathop{S}_{t=1}^{t=v} (\alpha_{st}) = \frac{n_{s.}}{N} \quad \text{and} \quad q_t = \frac{n_{.t}}{N}.$$

Thus the "best" form to give the parent population margins, i.e. that which makes χ^2 a minimum for this sample, are frequencies proportional to those of the sample itself. In this case

$$\phi^2 = \frac{1}{N} S \frac{\left(n_{st} - \frac{n_{s.} n_{.t}}{N}\right)^2}{\frac{n_{s.} n_{.t}}{N}} = \frac{\chi^2}{N} \quad \text{.........................(x).}$$

Now in order that (ix) may hold, the a_{st} series must be supposed constant throughout the sampling, i.e. we are to suppose $n_{s.}$ and $n_{.t}$ remain the same for all further samples and only n_{st} to vary. Thus the successive samples cannot be arranged as contingency tables. For if we change $n_{s.}$ and $n_{.t}$ with each sample we are making a new parent population with each sample, and the samples cannot then be supposed drawn from the *same* parent population*.

We now reach a case of which much use has been made, but which I think needs very careful handling.

As in the previous section we have

$$N(1 + \phi^2) = \chi^2 + N = S\left(\frac{n_{st}^2}{N a_{st}}\right) \quad \text{.......................(xi)}$$

and we desire to determine whether there is independence in certain results, which are assumed (as hypothesis) to be sampled from a population of zero contingency.

We can best illustrate this by an example given by Dr R. A. Fisher†, dealing with Wachter's data for back-crosses in mice. He gives a fourfold table running as follows:

TABLE II.

		Black Self	Black Piebald	Brown Self	Brown Piebald	Totals
Coupling	F_1 Males	88	82	75	60	305
	F_1 Females	38	34	30	21	123
Repulsion	F_1 Males	115	93	80	130	418
	F_1 Females	96	88	95	79	358
Totals		337	297	280	290	1204

Now neglecting the marginal totals the problem seems to be: Could the 16 cell frequencies have arisen from sampling a parent population with no contingency?—Thus $a_{st} = p_s \times q_t$. But we have taken of the four categories four samples of the sizes 305, 123, 418, 358. Are we going to confine our attention to such distributions

* There is nothing to prevent ϕ^2 with $n_{s.}$ and $n_{.t}$ varying from sample to sample being used as a statistical coefficient, but in that case χ^2 is not $N\phi^2$, if we mean by χ^2 that which is distributed by the law of equations (ii) or (ix). Result (x) seems to justify the usual expression for ϕ^2 as a measure of the departure from independence, ($n_{st} = n_{s.} n_{.t}/N$, or $\chi^2_{\min} = 0$), when we have no knowledge of the parent population.

† *Statistical Methods for Research Workers*, 3rd ed. p. 86.

as arise when we repeatedly make samples of these same sizes? Apparently we are to do so and this though somewhat artificial could be carried out, if with difficulty. We thus reach what I have termed a coefficient of partial contingency* with four linear equations of condition among the n_{st}'s. This will reduce the n' of our (χ^2, P) table by *three*, as that table takes account of the total size of the sample 1204. So far so good. But now we come to the horizontal marginal totals. These must vary from experiment to experiment; what reason is there for treating $\frac{337}{1204}$, $\frac{297}{1204}$, $\frac{280}{1204}$ and $\frac{290}{1204}$ as the values of q_1, q_2, q_3 and q_4 in the hypothetical parent population of no contingency from which we suppose our four samples extracted? Out of all possible repetitions of the four series of crossings, those giving the horizontal marginal totals coinciding with those of these actually performed experiments in the several categories of mice must be of the highest rarity and we should find it practically impossible to obtain such sets. It would seem that in choosing the horizontal marginal totals as the values of the q_t's we are really repeating what was done in the biserial table, i.e. assuming that as we do not know the q_t's we shall do the "best" we can by supposing they agree with those of the observed frequencies. If we assume that all further experiments are to give these same marginal frequencies, we again limit by three more linear relations our contingency, or, in looking up P from χ^2 we must reduce n' by six, or enter the table with $n' = 10$. The problem we are then answering is this: If we made further quadruple experiments each having the same number of mice from each form of crossing, and each quadruple experiment giving precisely the same numbers of Black Self, Black Piebald, Brown Self, and Brown Piebald mice, in how many cases might (on the basis of independence) at least as great a value of χ^2 be expected? But this further limitation is unnecessary, if all we have assumed is a system of likely values for the q_t's, and suppose successive further samples not to give the same horizontal marginal totals. Dr Fisher appears to prefer the extreme limitation to 9 degrees of freedom. This forces his further samples into a partial contingency table form, with all the marginal frequencies identical with those of the actual sample.

Not only will the χ^2 and therefore the P, as I have shewn†, be dependent on the number of mice resulting from each cross, for example, be altered if we had 209 instead of 418 "Repulsion, F_1 Males," but in practice the repetition of the coat colour distribution in further quadruple experiments is unattainable. It seems awkward in applied science to say something will occur, if so and so be done, when the doing of the latter is practically impossible.

Some, if not all, the difficulty may be surmounted, if we turn back to our value of χ^2, namely:

$$\chi^2 + N = S\left(\frac{n_{st}^2}{Na_{st}}\right) = S\left(\frac{n_{st}^2}{Np_s q_t}\right),$$

* "On the General Theory of Multiple Contingency, with Special Reference to Partial Contingency," *Biometrika*, Vol. XI. pp. 145—158; see in particular p. 146.

† *Biometrika*, Vol. XXIV. pp. 302-303, footnote.

since the parent population is assumed by hypothesis to have independence—and ask, supposing the p_s's to be fixed, what are the "best" values to give to the q_t's on the basis of the observed results?—Our answer is as before that the greatest probability for the observed results on the basis of an independent parent population will be obtained by choosing the q_t's so as to make χ^2 a minimum.

We proceed to make

$$\chi^2 = \underset{s}{S}\ \underset{t}{S}\left(\frac{n_{st}^2}{Np_s q_t}\right) - N \quad \text{(xii)}$$

a minimum subject to the condition that

$$\underset{t}{S}(q_t) = 1.$$

We have with λ a indeterminate multiplier

$$\underset{s}{S}\left(\frac{n_{st}^2}{Np_s}\right)\frac{1}{q_t^2} = \lambda,$$

or multiplying up by q_t and summing for t:

$$\chi^2_{\text{min.}} + N = \underset{t}{S}(\lambda q_t) = \lambda.$$

Hence:

$$q_t^2 = \frac{\underset{s}{S}\left(\dfrac{n_{st}^2}{Np_s}\right)}{\chi^2_{\text{min.}} + N},$$

and it follows on substituting for q_t in (xii) that

$$\chi^2_{\text{min.}} = \frac{1}{N}\left[\underset{t}{S}\left\{\underset{s}{S}\left(\frac{n_{st}^2}{p_s}\right)\right\}^{\frac{1}{2}}\right]^2 - N \quad \text{(xiii)},$$

and that

$$q_t = \frac{\left\{\underset{s}{S}\left(\dfrac{n_{st}^2}{p_s}\right)\right\}^{\frac{1}{2}}}{\underset{t}{S}\left\{\underset{s}{S}\left(\dfrac{n_{st}^2}{p_s}\right)\right\}^{\frac{1}{2}}} \quad \text{(xiv)},$$

or q_t is always < 1 and $\underset{t}{S}(q_t) = 1$.

Applying these results to our example, we proceed first to square all the terms n_{st} and take the reciprocals of the p_s's. Thus we find proceeding down a column $\underset{s}{S}\left(\frac{n_{st}^2}{p_s}\right)$ and then take the square roots of these expressions. We obtain:

$$\sqrt{\underset{s}{S}\left(\frac{n_{s1}^2}{p_s}\right)} = 337{\cdot}3310, \quad \sqrt{\underset{s}{S}\left(\frac{n_{s2}^2}{p_s}\right)} = 298{\cdot}0192,$$

$$\sqrt{\underset{s}{S}\left(\frac{n_{s3}^2}{p_s}\right)} = 282{\cdot}4913, \quad \sqrt{\underset{s}{S}\left(\frac{n_{s4}^2}{p_s}\right)} = 296{\cdot}9776,$$

and accordingly:

$$\underset{t}{S}\sqrt{\underset{s}{S}\left(\frac{n_{st}^2}{p_s}\right)} = 1214{\cdot}8191,$$

$$\chi^2_{\text{min.}} = \frac{(1214{\cdot}8191)^2}{1204} - 1204 = 21{\cdot}7354,$$

and:

$$q_1 = {\cdot}277{,}6800, \quad q_2 = {\cdot}245{,}3198, \quad q_3 = {\cdot}232{,}5378, \quad q_4 = {\cdot}244{,}4624.$$

Dr Fisher obtains for χ^2 the value

$$\chi_0^2 = 21{\cdot}832,$$

working his table as an ordinary contingency table and his q_t's, which are his horizontal marginal totals divided by his $N(=1204)$, will be

$$q_1 = {\cdot}279{,}9003, \quad q_2 = {\cdot}246{,}6777, \quad q_3 = {\cdot}232{,}5582, \quad q_4 = {\cdot}240{,}8638.$$

The reader will say: "It is true your χ^2 is less than Dr Fisher's, but your q_t's are so close to his and both your χ^2's differ only by an insignificant difference, that it is not clear why an attempt should be made to improve on them*." But there is really a wide difference between the two methods of approach! Suppose we knew or guessed the p_s's and q_t's of the parent population. Then we should use Equation (xii) to compute χ^2 and we should enter the (χ^2, P) table with $n' = 16$, because there would not (beside the size of the sample) be any restrictions whatever on the freedom of our sample. By fixing the p_s's, because there is no "natural" size for the relative numbers of matings we may make artificially, we have reduced our conclusions, whatever they may be, to apply only to a repetition of experiments of these sizes. We have destroyed the possibility of a general law; we cannot assert that for other sizes of samples, we should deduce the same conclusion. We have reached a conclusion for a narrower universe by sacrificing three degrees of freedom.

But at any rate let us attempt to reach a conclusion for a broader universe by not sacrificing further degrees of freedom! If we make the coat-colour distributions to be the same for every set of quadruple types of matings, our conclusions will apply only to an absolutely restricted and practically irreproducible universe! But how can we avoid this? Only by assuming some set of q_t's for the hypothetical parent population. How may we do this? There are two obvious ways: (*a*) assume that the experiments give a good approximation to the required q_t's, or (*b*) find the most likely q_t's by the method of this paper, i.e. those which make the probability of the observed result a maximum. In either case we do not further restrict the degrees of freedom. Further quadruple samples will not have the horizontal marginal totals the same as those of the observed sample—i.e. will not take the form of contingency tables. What then has Dr Fisher done, when he reduces his degrees of freedom by still a further three? He has picked out of all the possible samples those which have their distributions the same for the coat colours. His conclusions therefore only hold for that extraordinarily limited universe.

It is of interest to note that in this particular case—it is far from being so in every case—the observed values of the coat-colour distribution are strikingly like the "best" values and lead to the same value practically of χ^2*.

If we took out the P that corresponds to that χ^2 with $n' = 16 - 3 = 13$, we find from (*a*) $P = {\cdot}040$, and from (*b*) $P = {\cdot}041$.

If we limit our n' to $10 = 16 - 6$, we find $P = {\cdot}010$. Now what does this signify? It denotes that with the narrower proposition when we experiment in such a manner

* The reason for the closeness of the two sets of q_t's is that our sample being large the horizontal margin total distribution gives nearly the same series of q_t's as the set which produces the minimum value of χ^2.

as always to get the same relative numbers of coat colours(!) we may predict that the four series ($P = \cdot 01$) are not homogeneous, whereas we are far less certain of this ($P = \cdot 04$) when we take experiments which could be fairly easily repeated. But in the former case we do not know whether the departure from independence really lies in genetic conditions, or is due to the restraints which have been put on the distribution of coat colour. The effect of abolishing these restraints appears at least to suggest that they have contributed to the result.

Of course the desirable thing would be to abolish all restraints except the total size of the sample, but this is impossible with regard to the p_s's, for their arbitrary character lies in the very nature of the experiment. To show to what extent the arbitrary choice of p_s's effects our conclusions, I will take the following table, which is obtained practically by doubling the number of "Coupling, F_1 Males" and halving the number of "Repulsion, F_1 Males." There appears to be nothing more arbitrary in this than in the results of the observed mating type proportions.

TABLE III.

		Black Self	Black Piebald	Brown Self	Brown Piebald	Totals
Coupling	F_1 Males	176	164	150	120	610
	F_1 Females	38	34	30	21	123
Repulsion	F_1 Males	58	46	40	65	209
	F_1 Females	96	88	95	79	358
Totals		368	332	315	285	1300

Following Dr Fisher, that is making the relative coat-colour distribution constrained, we obtain:

$$\chi^2 = 16\cdot 2710 \quad \text{and} \quad P = \cdot 062.$$

Thus even with the $P = \cdot 05$ limit, we could not now assert that the observed departures from independence are not of a magnitude ascribable to chance.

This illustration points only too strongly to the caution requisite in applying this method to draw conclusions from observed data; the conclusion drawn will depend on the number of mice, and therefore on the relative number of crossings made in each one of the four sections of the quadruple experiment, and these are at the choice of the experimenter.

If we do not limit our judgment to experiments giving always the same relative proportions of coat colour, then we must enter with $n' = 13$, instead of 10. If we take q_t's of the parent population to be those given by the observed values, i.e.

$$q_1 = \cdot 283{,}0769, \quad q_2 = \cdot 253{,}3846, \quad q_3 = \cdot 242{,}3077, \quad q_4 = \cdot 291{,}2308,$$

we have $\chi^2 = 16\cdot 2710$, and $P = \cdot 180$, a value which quite prohibits our concluding that the departures from independence are not ascribable to chance. If we use the "best" values of the q_t's, they are:

$$q_1 = \cdot 281{,}5800, \quad q_2 = \cdot 254{,}4918, \quad q_3 = \cdot 241{,}9760, \quad q_4 = \cdot 221{,}9522,$$

again unusually close to the observed values.

We then have for the "best" χ^2:

$$\chi^2_{\text{min.}} = 16{\cdot}2114,$$

giving $P = {\cdot}182$, which is practically the same as we find from using the observed q_t's to represent the parent population.

I trust this discussion has to some extent cleared up the difficulties which await those who use a contingency table of multiple rows to question whether the multiple series involved in those rows may be treated as homogeneous, i.e. possible samples from a common parent population. We have seen that in Dr Fisher's approach to the problem two difficulties arise. The first from the arbitrary numbers of mice in each experimental series, and the second from the constraints enforced on the coat colours. Dr Fisher is really proposing a series of experiments, each individual experiment giving the same numbers of mice from each type of mating as occur in the observed experiment,—this may be needful,—and further the same relative proportions of coat colours. The latter is not needful, and would be practically impossible to achieve. Dr Fisher concludes that if he could repeat the multiple experiment under these conditions the χ^2 would correspond to a low probability, but there is no evidence that this result flows from the genetic constitution of the mice. Indeed, if we alter the numbers of mice from each type of mating, i.e. alter the number of matings, and leave in our experiments the distribution of coat colours to freely adjust themselves, we find χ^2 can be so modified as to provide a probability, which is far from suggesting that the departures from independence are not of a magnitude to be ascribed to chance. The method therefore needs great caution in use, and there should always be an exact statement of what the problem supposed to be answered really is.

It will be seen that the method of the contingency table fails in stating clearly what is the homogeneous population from which we are supposing the four series to be drawn, and, admitting the difficulty of the problem, I prefer to attack it by using (i) and comparing each pair of series with one another. The question is: What value shall we give to the p_s's for each pair? It seems very much better not to use the observed sum of the columns, but to adopt for p_s the value given in (v) and accordingly for χ^2 its minimum value in (iv). If we use these we have six pairs of series to compare, and owing to the simplicity of (iv), the arithmetical work is no more laborious than that of finding χ^2 from a 4×4 table.

We take the reciprocals of the marginal totals column and by aid of them determine the relative frequencies of each row. Thus we get the table:

TABLE IV.

Series	n_1/N	n_2/N	n_3/N	n_4/N	N
A	·288,5246	·268,8525	·245,9017	·196,7212	305
B	·308,9431	·276,4228	·243,9024	·170,7317	123
C	·275,1196	·222,4880	·191,3876	·311,0048	418
D	·268,1564	·245,8101	·265,3631	·220,6704	358

We now take the differences between the entries in each pair of rows, add these differences and squaring their sum multiply by the corresponding value of $\nu = NN'/(N+N')$, where N and N' are given in the last column.

For example, taking A and B, the differences, regardless of sign, are

$$\cdot 020{,}4185, \quad \cdot 007{,}5703, \quad \cdot 001{,}9993, \quad \cdot 025{,}9895,$$

giving the sum ·055,9776, and its square ·0031,3349,

$$\nu = 305 \times 123/(305+123) = 87{\cdot}6519,$$

and thus $$\chi^2 = 87{\cdot}6519 \times \cdot 0031{,}3349 = \cdot 2747.$$

The P of course is to be looked up under the number of categories, i.e. $n' = 4$.

Proceeding in this way, we find

$$\left.\begin{array}{lll} A \text{ and } B: & \chi^2 = 0{\cdot}2747, & P = \cdot 945 \\ A \text{ and } C: & \chi^2 = 9{\cdot}2122, & P = \cdot 027 \\ A \text{ and } D: & \chi^2 = 1{\cdot}2414, & P = \cdot 746 \\ B \text{ and } C: & \chi^2 = 7{\cdot}4791, & P = \cdot 059 \\ B \text{ and } D: & \chi^2 = 1{\cdot}8668, & P = \cdot 603 \\ C \text{ and } D: & \chi^2 = 7{\cdot}3023, & P = \cdot 064 \end{array}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(xv)}.$$

By the $P = \cdot 02$ criterion, none of these are significant; by the $P = \cdot 05$, the A and C differences are. But we see at once that while the series A, B and D might be considered as samples from the same population, the probability that any one of them and C can be considered as such is of a much lower order. Accordingly we take the sum of A, B and D and test it against C. Thus we have the relative frequencies:

Series	n_1/N	n_2/N	n_3/N	n_4/N	N
$A+B+D$	·282,4427	·259,5420	·254,4529	·203,5624	786
C	·275,1196	·222,4880	·191,3876	·311,0048	418

giving the differences regardless of sign:

$$\cdot 007{,}3231 \quad \cdot 037{,}0540 \quad \cdot 063{,}0653 \quad \cdot 107{,}4424$$

with a total of ·214,8848 and square ·0461,7548, this with $\nu = 272{\cdot}8804$ leads to $\chi^2 = 12{\cdot}6004$ and $P = \cdot 006$.

The advantages of this process are that Table IV enables us to make any analysis we please of the material as we proceed, and that while (xv) is not definite, it indicates the line on which we can get a perfectly definite result. Finally we are certain that whatever parent population we may take for a pair, that chosen is the one which makes the observed result most probable; in other words if definite heterogeneity may be predicted on the result thus obtained, it would certainly be predicted on any other assumption of a parental population distribution. We see that if $A+B+D$ and C be supposed to be two samples from the same parent population at a maximum whatever that population might be, two such samples could not arise more than 6 times in 1000 trials.

The method is straightforward, the arithmetic simple, and we take P out of our (χ^2, P) table with n' equal to the number of categories in the series.

This method of approaching the problem is of course not free, any more than the contingency table process, from the variation in χ^2—and therefore in P—produced by the artificial choice of numbers of matings and the resulting numbers of offspring. This difficulty is introduced by the factor $\nu = NN'/(N+N')$. Supposing we keep the relative percentages of coat colour the same in the two series as well as the total number of mice, the maximum value of ν, for $A+B+D$ as compared with C, will be $\frac{1}{4}(N+N') = 301$, in our case leading to $\chi^2 = 13{\cdot}8988$ and $P = {\cdot}003$, which makes some difference in P, but not in our conclusion*. The relative size of the samples appears in such a simple form when we proceed from the biserial method, that it is fairly easy to appreciate their influence. Failing any "natural" distribution of the totals in the sub-experiments, it would perhaps be the simplest rule for the research worker to keep them as near an equality as possible.

* The choice of the ratio $N:N'$ may easily make a difference in our conclusions. Thus if in A we had had 514 mice and in C 209, giving the same total 723, we should have had $\nu = 148{\cdot}58368$ instead of $176{\cdot}33472$, and, if the colour percentages in each series had remained much the same, we should have $\chi^2 = 7{\cdot}7625$ instead of $9{\cdot}2122$ and $P = {\cdot}188$ instead of the ${\cdot}027$ of (xv). The number of mice in the C group (Repulsion, F_1 Males) is the largest of all four sub-experiments, and we must be very cautious to allow for this when, on the basis of the χ^2 test, we attribute to C a genetic differentiation from A, B and D.

CERTAIN GENERALIZATIONS IN THE ANALYSIS OF VARIANCE*.

BY S. S. WILKS, PH.D., National Research Fellow in Mathematics, Columbia University, New York.

CONTENTS.

		PAGE
1.	Introductory Remarks	471
2.	Solutions of two integral Equations	473
3.	Generalization of the Variance of a Sample	476
4.	Moments and Distribution of the Ratio of independent generalized Variances	478
5.	Ratio of the independent Generalized Variance to any one of its principal Minors	480
6.	Generalization of the Correlation Ratio	482
7.	Generalization of $1-\eta^2$	485
8.	Generalization of "Student's" Ratio	487
9.	Generalization of the λ-Criterion appropriate to k Samples	488
10.	Moments and Distributions of Ratios of Determinants of Correlation Coefficients	491

1. *Introductory Remarks.* The theory of small samples has been developed to a large extent from problems involving a single variate. The extension of the theory to samples from multivariate populations has been made rather slowly and it is far from complete at present. It was not until 1928 that Wishart† found the simultaneous sampling distribution of the variances and covariances in samples from a multivariate normal universe, whereas Fisher‡ solved the problem for a bivariate normal population in 1915. In the same paper, Fisher found the distribution of the correlation coefficient and in 1928 he solved the corresponding problem for the multiple correlation coefficient§. The distribution introduced by "Student"‖ in 1908 in his analysis of the ratio of the deviation of the mean of a sample from that of the population to the standard deviation of the sample was more rigorously obtained by Fisher¶ in 1925. At the same time Fisher extended its application to sampling fluctuations of regression coefficients, differences of means and other problems which

* Presented to the *American Mathematical Society*, March 26, 1932.

† J. Wishart: "The generalized Product Moment Distribution in Samples from a normal multivariate Population," *Biometrika*, Vol. xxA. (1928), pp. 32—52.

‡ R. A. Fisher: "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an indefinitely large Population," *Biometrika*, Vol. x. (1915), pp. 507—521.

§ R. A. Fisher: "The general sampling Distribution of the Multiple Correlation Coefficient." *Proceedings of the Royal Society of London*, Series A, Vol. cxxi. (1928), pp. 654—678.

‖ "Student": "The probable Error of a Mean," *Biometrika*, Vol. vi. (1908—1909), pp. 1—25.

¶ R. A. Fisher: "Applications of 'Student's' distribution," *Metron*, Vol. v. No. 3 (1925), pp. 90—104.

involve essentially a single variate. These ideas were generalized in 1931 by Hotelling* who found the distribution of a quantity T which, when divided by the square root of the number of degrees of freedom, is a natural extension of "Student's" original z to a sample from a multivariate normal population. We find very few additional extensions of the kind with which we are concerned existing in the literature.

Statistical coefficients which have not been adequately generalized for samples from a multivariate population include the variance, ratio of variances, correlation ratio, standard error of estimate when all variates are drawn at random, and certain maximum likelihood test criteria developed for one-variable problems by Pearson and Neyman†. As early as 1876 Helmert‡ found the distribution of the sum of squares of deviations of a set of normally and independently distributed quantities from the population mean, and in 1900 Karl Pearson§ solved the same problem for the case where there is correlation among the variates and found the distribution of χ^2. In 1908 "Student"|| suggested the form of the distribution of the sum of squares of the deviations of the variates of a sample from the sample mean, which was verified in 1915 by Fisher¶. By means of the distribution of the ratio of two independently distributed variances, Fisher** has found the distribution of the multiple correlation coefficient and the correlation ratio in samples from normal populations in which these quantities are zero. He has extended the use of this distribution to the problem of testing for the significance of variations in certain subvariances into which the variance of a sample can be analyzed and has developed the theory of intraclass correlations. Romanovsky†† introduced an extension of the ratio of variances and found the sampling distribution of a quantity H which is the average of the ratios of variances for two samples from a multivariate population. But this does not seem to be a perfectly natural extension of the variance problem for samples from multivariate populations as we shall see later. In 1928 E. S. Pearson and

* H. Hotelling: "The Generalization of 'Student's' Ratio," *Annals of Mathematical Statistics*, Vol. II. (1931), pp. 359—378.

† J. Neyman and E. S. Pearson: "On the Use and Interpretation of certain test Criteria for purposes of statistical Inference," *Biometrika*, Parts I. and II. Vol. XX^A. pp. 175—240, 263—294.

J. Neyman and E. S. Pearson: "On the Problem of k Samples," *Bulletin de l'Académie Polonaise des Sciences et des Lettres, Série A, Sciences mathématiques*, 1931, pp. 460—481.

‡ C. F. Helmert: "Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und über einige damit in Zusammenhang stehende Fragen," *Zeitschrift für Mathematik und Physik*, Vol. XXI. (1876), pp. 192—219.

§ K. Pearson: "On the Criterion that a given set of Deviations from the probable in the case of correlated Variables is such that can be reasonably supposed to have arisen from Random Sampling," *Philosophical Magazine*, 5th series, Vol. L. (1900), pp. 157—175.

|| "Student": *loc. cit.* [Helmert also in 1876 proved the equation for the distribution of the sums of the squares of the deviations about the *sample* mean. See *Astronomische Nachrichten*, Bd. LXXXVIII. S. 122, or *Biometrika*, Vol. XXIII. pp. 416—418. Ed.]

¶ R. A. Fisher: *Biometrika*, Vol. X. (1915), p. 507.

** R. A. Fisher: "On a Distribution yielding the Error Functions of several well-known Statistics," *Proceedings of the International Mathematical Congress*, Toronto (1924), Vol. II. pp. 805—813.

†† V. Romanovsky: "On the Criteria that two given Samples belong to the same Normal Population," *Metron*, Vol. VII. No. 3 (1928), pp. 3—46.

Neyman* began a series of papers in which they adopted the principle of maximum likelihood as a means of obtaining criteria for testing various hypotheses in statistical inference. Among others they have obtained criteria appropriate to the hypotheses that two or more samples are drawn from the same normal population; that a sample is drawn from a population with a specified mean; that two or more samples from populations having identical variances come from populations with identical means and a similar hypothesis stated by interchanging variances and means. They have thus far confined their work primarily to samples from normal populations of a single variable.

Investigators dealing with samples of two or more correlated variables are confronted with the need of extended forms of the above statistical mechanisms. For example, measurements of several anthropological characters are obtained on two or more groups of men; how can we test the hypothesis that they are from the same race by a consideration of their variances and covariances? Similar problems arise in psychology concerning certain mental tendencies of two or more groups of individals who have been measured on the basis of several mental traits; and so on for other fields of statistical investigation.

In this paper it is the purpose of the author to find the moments and distributions of some of the foregoing statistical coefficients generalized for samples from a multivariate normal population and to exhibit a method of attack which seems to be novel in its application. Another problem which will be considered concerns the moments and distributions of the determinants and certain ratios of determinants of correlation coefficients in multivariate samples, from which a certain generalization of the multiple correlation coefficient is obtained.

2. *Solutions of two integral Equations.* The moments of the class of statistical coefficients which we shall consider are of a form which is a constant multiple of a ratio of products of gamma functions. Most of the distributions can be derived from the solutions of two types of integral equations. We shall desigrate these two types by (A) and (B) and find their solutions before considering the main part of the problem.

Type A. The first to be considered is of the form

$$\int_0^\infty z^k f(z)\,dz = B^k \frac{\Gamma(a_1+k)\,\Gamma(a_2+k)\dots\Gamma(a_n+k)}{\Gamma(a_1)\,\Gamma(a_2)\dots\Gamma(a_n)} \quad \dots\dots\dots\dots(\text{A}),$$

where k and the a's are real and positive and B and $f(z)$ are independent of k.

By definition,
$$\Gamma(a_i+k) = \int_0^\infty \theta_i^{a_i+k-1} e^{-\theta_i} d\theta_i.$$

Hence, as far as the moments are concerned, the problem of finding $f(z)$ is equivalent

* J. Neyman and E. S. Pearson: *Biometrika*, Vol. xx^A. Parts I. and II. pp. 175—240.

J. Neyman and E. S. Pearson: *Bulletin de l'Académie Polonaise des Sciences et des Lettres, Série A, Sciences mathématiques*, 1930 and 1931.

to that of finding the distribution of the product $z = B\theta_1\theta_2 \dots \theta_n$, where θ_i has the distribution

$$\frac{1}{\Gamma(a_i)} \theta_i^{a_i - 1} e^{-\theta_i} d\theta_i, \qquad (i = 1, 2, \dots n).$$

Letting $\theta_n = \dfrac{z}{B\theta_1\theta_2 \dots \theta_{n-1}}$, and substituting in

$$\prod_{i=1}^{n} \frac{1}{\Gamma(a_i)} \theta_i^{a_i-1} e^{-\theta_i} d\theta_i,$$

we have for the distribution of z,

$$f(z) = \frac{B^{-a_n} z^{a_n - 1}}{\Gamma(a_1)\Gamma(a_2)\dots\Gamma(a_n)} \int_0^\infty \int_0^\infty \dots \int_0^\infty \theta_1^{a_1 - a_n - 1} \theta_2^{a_2 - a_n - 1} \dots \theta_{n-1}^{a_{n-1} - a_n - 1}$$

$$\times e^{-\theta_1 - \theta_2 - \dots - \theta_{n-1} - \frac{z}{B\theta_1 \dots \theta_{n-1}}} d\theta_1 d\theta_2 \dots d\theta_{n-1} \quad \dots\dots(1).$$

By making the transformation $\theta_1\theta_2 \dots \theta_i = v_i$ $(i = 1, 2, \dots n-1)$, we can write

$$f(z) = \frac{B^{-a_n} z^{a_n - 1}}{\Gamma(a_1)\Gamma(a_2)\dots\Gamma(a_n)} \int_0^\infty \int_0^\infty \dots \int_0^\infty v_1^{a_1 - a_2 - 1} v_2^{a_2 - a_3 - 1} \dots v_{n-1}^{a_{n-1} - a_n - 1}$$

$$\times e^{-v_1 - \frac{v_2}{v_1} - \dots - \frac{v_{n-1}}{v_{n-2}} - \frac{z}{Bv_{n-1}}} dv_1 dv_2 \dots dv_{n-1} \quad \dots\dots\dots(2).$$

The author has succeeded in integrating this expression only for special values of the a's and small values of n, which will be considered later.

We note that the integral in (A) exists for all positive values of k and hence all functions satisfying the integral equation (A) must have their kth moments identical ($k = 0, 1, 2, \dots$). The uniqueness of the continuous solution (2) can be established by the use of Stekloff's* application of the theory of closure to the problem of moments.

Since we are dealing with non-negative functions, it is to be noted that if we had not known the range of z in (A), it could be argued from Stekloff's theory that it must be from 0 to ∞. This type of argument is especially important in establishing the range of statistical coefficients which we shall consider.

Type B. Next we shall consider the integral equation

$$\int_0^B w^k g(w)\, dw = CB^k \frac{\Gamma(b_1 + k)\Gamma(b_2 + k)\dots\Gamma(b_n + k)}{\Gamma(c_1 + k)\Gamma(c_2 + k)\dots\Gamma(c_n + k)} \quad \dots\dots\dots\dots(B),$$

where $C = \dfrac{\Gamma(c_1)\Gamma(c_2)\dots\Gamma(c_n)}{\Gamma(b_1)\Gamma(b_2)\dots\Gamma(b_n)}$ and B and $g(w)$ are independent of k; where also the b's and c's are two sets of real and positive numbers such that there exists at least one way of pairing them such that each b is less than its corresponding c. Thus we assume without loss of generality that $b_i < c_i$, $(i = 1, 2, \dots n)$.

* W. Stekloff: "Quelques applications nouvelles de la théorie de fermeture au problème de représentation approchée des fonctions et au problème des moments," *Mémoire de l'Académie Impériale des Sciences de St Pétersbourg*, Vol. XXXII. No. 4, 1914.

Let us multiply and divide the expression on the right of (B) by $\prod_{i=1}^{n}(c_i - b_i)$. Then, since

$$\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^1 t^{p-1}(1-t)^{q-1}\,dt,$$

our problem is resolved to that of finding the distribution of the quantity $w = Bt_1t_2\ldots t_n$ where the simultaneous distribution of the t's is given by

$$\prod_{i=1}^{n}\frac{\Gamma(c_i)}{\Gamma(b_i)\Gamma(c_i-b_i)}t_i^{b_i-1}(1-t_i)^{c_i-b_i-1}\,dt_i \qquad\ldots\ldots\ldots\ldots(3).$$

If we make the substitution

$$t_n = \frac{w}{Bt_1t_2\ldots t_{n-1}}$$

in (3), we have

$$g(w) = K\frac{w^{b_n-1}}{B^{b_n}}\int_{L_1}^1\int_{L_2}^1\ldots\int_{L_{n-1}}^1 t_1^{b_1-b_n-1}t_2^{b_2-b_n-1}\ldots t_{n-1}^{b_{n-1}-b_n-1}$$
$$\times(1-t_1)^{c_1-b_1-1}(1-t_2)^{c_2-b_2-1}\ldots(1-t_{n-1})^{c_{n-1}-b_{n-1}-1}\left(1-\frac{w}{Bt_1\ldots t_{n-1}}\right)^{c_n-b_n-1}dt_1\ldots dt_{n-1}$$
$$\ldots\ldots\ldots(4),$$

where $\quad K = \prod_{i=1}^{n}\frac{\Gamma(c_i)}{\Gamma(b_i)\Gamma(c_i-b_i)},\quad L_i = \frac{w}{Bt_1t_2\ldots t_{i-1}},\quad t_0 = 1,\quad (i = 1, 2, \ldots n-1).$

In order to make the limits of the integrations in (4) independent of the variables, we shall make the following transformation

$$t_i = 1 - v_i\left(1-\frac{w}{Bt_1t_2\ldots t_{i-1}}\right),\qquad (i = 1, 2, \ldots n-1).$$

As the result, we obtain

$$g(w) = \frac{Kw^{b_n-1}\left(1-\frac{w}{B}\right)^{\gamma_n-\beta_n-1}}{B^{b_n}}\int_0^1\int_0^1\ldots\int_0^1 v_1^{c_1-b_1-1}v_2^{c_2-b_2-1}\ldots v_{n-1}^{c_{n-1}-b_{n-1}-1}$$
$$\times(1-v_1)^{\gamma_{n-1}-\beta_{n-1}-1}(1-v_2)^{\gamma_{n-2}-\beta_{n-2}-1}\ldots(1-v_{n-1})^{\gamma_1-\beta_1-1}$$
$$\times\left[1-v_1\left(1-\frac{w}{B}\right)\right]^{b_1-c_2}\left[1-\{v_1+v_2.(1-v_1)\}\left(1-\frac{w}{B}\right)\right]^{b_2-c_3}\ldots$$
$$\times\left[1-\{v_1+v_2.(1-v_1)+\ldots+v_{n-1}.(1-v_1).(1-v_2)\ldots(1-v_{n-2})\}\left(1-\frac{w}{B}\right)\right]^{b_{n-1}-c_n}$$
$$\times\,dv_1dv_2\ldots dv_{n-1}\ldots\ldots\ldots(5),$$

where $$\gamma_i = \sum_{j=0}^{i-1}c_{n-j},\quad \beta_i = \sum_{j=0}^{i-1}b_{n-j},\qquad (i = 1, 2, \ldots n).$$

Since the distribution of w has the range 0 to B, we have, for $B \geqslant w > 0$, $1-\frac{w}{B} < 1$. Then we can show by induction that

$$\{v_1+v_2.(1-v_1)+\ldots+v_i.(1-v_1).(1-v_2)\ldots(1-v_{i-1})\}\left(1-\frac{w}{B}\right) < 1\quad\ldots(6)$$

for $0 \leqslant v_i \leqslant 1$, $(i = 1, 2, \ldots n-1)$. Therefore, the series which results from expanding all of the factors in (5) involving the term (6) is uniformly convergent in the v's over the field of integration and can be integrated term by term. This process yields the distribution function $g(w)$ which, again, is unique.

3. *Generalization of the Variance of a Sample.* Wishart* has shown that the simultaneous distribution of the variances and covariances of a sample of N items from an n-variate normal population is given by

$$\frac{|A_{ij}|^{\frac{N-1}{2}}}{(\pi)^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{N-i}{2}\right)} e^{-\sum_{i,j=1}^{n} A_{ij} a_{ij}} |a_{ij}|^{\frac{N-n-2}{2}} da \quad \ldots\ldots\ldots\ldots\ldots(7),$$

where $|A_{ij}|$ is the nth order determinant of the elements $A_{ij} = \frac{N\Delta_{ij}}{2\sigma_i\sigma_j\Delta}$. Δ_{ij} denotes the co-factor corresponding to ρ_{ij} in the determinant $\Delta = |\rho_{ij}|$ of population correlations and σ_i the standard deviation in the population of the ith variate. Thus, if $\lambda = |\sigma_i\sigma_j\rho_{ij}|$ is defined as the generalized variance of the population, then $|A_{ij}| = N^n [2^n\lambda]$. da is the product of the differentials of all of the a's. The elements of the determinant $|a_{ij}|$ are the variances and covariances from the sample defined as

$$a_{ij} = a_{ji} = \frac{1}{N} \sum_{\alpha=1}^{N} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j), \qquad (i, j = 1, 2, \ldots n),$$

where $\bar{x}_i = \frac{1}{N} \sum_{\alpha=1}^{N} x_{i\alpha}$ is the sample mean of the ith variate, and $x_{i\alpha}$ is the value of the ith variate x_i, for the αth individual.

We shall adopt as the generalized sample variance the determinant $|a_{ij}|$. This quantity for n-variate samples and the ordinary variance for samples of one variate are similar, not only in the manner in which they enter the distribution of their component parts (there being only one part in the case of single variate samples), but in the way they arise in maximizing likelihood functions†. For example, the maximum of the likelihood expressed by (7) for variations of the population parameters A_{ij} is $C|a_{ij}|^{-\frac{n+1}{2}}$, and if the L-function of the means is taken into account, the maximum of the joint L-function is proportional to $|a_{ij}|^{-\frac{n+2}{2}}$. In one-variable samples, the maximum L-functions for the two cases are proportional to a^{-1} and $a^{-\frac{3}{2}}$ respectively, where a is the ordinary variance.

Let us denote $|a_{ij}|$ by ξ, and proceed to find its kth moment $M_k(\xi)$. From the fact that the integral of (7) over the field of possible values of the a's is unity, we have (using abbreviated notation)

$$\int e^{-\sum_{i,j=1}^{n} A_{ij} a_{ij}} |a_{ij}|^{\frac{N-n-2}{2}} da = \frac{\pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{N-i}{2}\right)}{|A_{ij}|^{\frac{N-1}{2}}} \quad \ldots\ldots\ldots\ldots\ldots(8).$$

* J. Wishart: *loc. cit.*

† In this paper we are primarily interested in various functions of the means, variances and covariances of a sample from a normal population. For this reason we shall express the probability of a sample in terms of the probability function F of its means, variances and covariances rather than in terms of the probabilities of the individual items of the sample. F is, of course, a function of the means, variances and covariances of both the sample and the population and is the product of two independent functions F_a and F_m, where F_a is the distribution function of the variances and covariances and F_m is that of the means. For a specified sample (i) F, (ii) F_a and (iii) F_m may be considered as functions of population parameters, and will be called likelihood functions or simply L-functions of (i) the sample, (ii) the variances and covariances, and (iii) the means.

Then $$M_k(\xi)=\frac{1}{G}\int e^{-\sum_{i,j=1}^{n}A_{ij}a_{ij}}\,|a_{ij}|^{\frac{N+2k-n-2}{2}}\,da \quad\text{.....................(9)},$$

where G is the expression on the right of (8). We have at once the value of the integral in (9) by substituting $N+2k$ for N in the right side of (8). Therefore

$$M_k(\xi)=\frac{A^{-k}\,\Gamma\left(\frac{N-1}{2}+k\right)\Gamma\left(\frac{N-2}{2}+k\right)\dots\Gamma\left(\frac{N-n}{2}+k\right)}{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-2}{2}\right)\dots\Gamma\left(\frac{N-n}{2}\right)} \quad\text{...(10)},$$

where $$A=|A_{ij}|=\frac{N^n}{2^n\sigma_1^2\sigma_2^2\dots\sigma_n^2\Delta}.$$

The fact that a factor N is concealed in each A_{ij} does not invalidate (10), because (8) holds for all positive values of σ_i and hence it holds when σ_i is replaced by

$$\sigma_i\sqrt{\frac{N+2k}{N}}.$$

Clearly, this process will absorb the increment $2k$ in the N multiplier of each A_{ij} and will not affect the increment of N entering at any other place in (8). The same result can be achieved by transforming the A's and a's by letting $A_{ij}=NB_{ij}$ and $a_{ij}=b_{ij}/N$ in (8), and finding the kth moment of $|b_{ij}|$ which is easily found to be $M_k(\xi)$.

If we denote the distribution of ξ by $D(\xi)$, we must have

$$\int \xi^k D(\xi)\,d\xi=M_k(\xi) \quad\text{..............................(11)},$$

an integral equation of type (A).

Therefore

$$D(\xi)=\frac{A^{\frac{N-n}{2}}\xi^{\frac{N-n-2}{2}}}{\prod_{i=2}^{n}\Gamma\left(\frac{N-i}{2}\right)}\int_0^\infty\int_0^\infty\dots\int_0^\infty (v_1v_2\dots v_{n-1})^{-\frac{1}{2}}\,e^{-v_1-\frac{v_2}{v_1}-\dots-\frac{v_{n-1}}{v_{n-2}}-\frac{A\xi}{v_{n-1}}}\,dv_1dv_2\dots dv_{n-1} \quad\text{......(12)},$$

and the range of ξ is from 0 to ∞.

If $n=1$, we get the well-known distribution of the variance in samples of a single variate

$$D_1(\xi)=\frac{\left(\frac{N}{2\sigma^2}\right)^{\frac{N-1}{2}}}{\Gamma\left(\frac{N-1}{2}\right)}\xi^{\frac{N-3}{2}}e^{-\frac{N}{2\sigma^2}\xi} \quad\text{........................(12 a)}.$$

For $n=2$*

$$D_2(\xi)=\frac{\sqrt{\pi}A_2^{\frac{N-2}{2}}\xi^{\frac{N-4}{2}}}{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-2}{2}\right)}e^{-2\sqrt{A_2\xi}}=\frac{2^{N-3}A_2^{\frac{N-2}{2}}\xi^{\frac{N-4}{2}}}{\Gamma(N-2)}e^{-2\sqrt{A_2\xi}} \quad\text{......(12 b)},$$

* If s_1 and s_2 were two sample standard deviations and r_{12} the correlation coefficient, then in this case $\xi=s_1^2s_2^2(1-r_{12}^2)$.

where $$A_2 = \frac{N^2}{4\sigma_1^2\sigma_2^2(1-\rho_{12}^2)}.$$

The author has thus far been unable to obtain $D(\xi)$ explicitly for larger values of n.

In practical work it may be desirable to use the nth root of the generalized variance, which would be the geometric mean of the variances of the n variates multiplied by the nth root of the determinant of correlations among the n variables. In this case the kth moments can be found from (10) by substituting $\frac{k}{n}$ for k. The distribution of the square root of the generalized variance for bivariate samples can be found from (12 b) by setting $\xi = t^2$, thus obtaining

$$g(t) = \frac{2^{N-2} A_2^{\frac{N-2}{2}} t^{N-3}}{\Gamma(N-2)} e^{-2\sqrt{A_2}t} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(12\,c).$$

Again, it may be important in certain situations to take the $2n$th root of the generalized variance, which would be the geometric mean of the standard deviations of the n variates multiplied by the $2n$th root of the determinant of correlations. In this case the moments can be found by substituting $\frac{k}{2n}$ for k.

4. *Moments and Distribution of the Ratio of independent generalized Variances.* The statistical coefficient to be considered here is a generalization of the ratio of two independently distributed variances whose distribution in samples of a single variate is used extensively by Fisher* in his analysis of variance.

Let us denote by ξ and η the generalized variances in two samples from n-variate populations in which the generalized variances are α and β respectively. If we let $\frac{\xi}{\eta} = \psi$, then since ξ and η are independent, the kth moment of ψ can be deduced from (10) as

$$M_k(\psi) = M_k(\xi)\,M_k(\eta) = \left(\frac{B}{A}\right)^k \prod_{i=1}^{n} \left[\frac{\Gamma\left(\frac{M-i}{2}+k\right)\Gamma\left(\frac{N-i}{2}-k\right)}{\Gamma\left(\frac{M-i}{2}\right)\Gamma\left(\frac{N-i}{2}\right)}\right] \ldots(13),$$

where $A = \frac{M^n}{2^n\alpha}$, $B = \frac{N^n}{2^n\beta}$, M and N being the numbers of individuals in the samples.

The distribution of ψ can be readily derived from the distributions of ξ and η, using the form given by (1). Accordingly, we have as the joint distribution of ξ and η,

$$KA^{\frac{M-n}{2}} B^{\frac{N-n}{2}} \xi^{\frac{M-n-2}{2}} \eta^{\frac{N-n-2}{2}} \int_0^\infty \int_0^\infty \ldots \int_0^\infty (t_1\theta_1)^{\frac{n-1}{2}-1} (t_2\theta_2)^{\frac{n-2}{2}-1} \ldots (t_{n-1}\theta_{n-1})^{\frac{1}{2}-1}$$
$$\times e^{-(t_1+\theta_1)-(t_2+\theta_2)-\ldots-(t_{n-1}+\theta_{n-1})-\frac{A\xi}{t_1t_2\ldots t_{n-1}}-\frac{B\eta}{\theta_1\theta_2\ldots\theta_{n-1}}} dT\,d\Theta \ldots(14),$$

* R. A. Fisher: *Proceedings of the International Mathematical Congress*, Vol. II. Toronto (1924).

where

$$K = \frac{1}{\prod_{i=1}^{n} \left[\Gamma\left(\frac{M-i}{2}\right)\Gamma\left(\frac{N-i}{2}\right)\right]}, \quad dT = dt_1 dt_2 \dots dt_{n-1}, \quad d\Theta = d\theta_1 d\theta_2 \dots d\theta_{n-1}.$$

Making the transformation

$$\frac{\xi}{\eta} = \psi, \quad t_i = \frac{s_i}{1-s_i}\theta_i, \qquad (i = 1, 2, \dots n-1),$$

and integrating with respect to the θ's and η, we can write the distribution of ψ in the following form:

$$F(\psi) = H\int_0^1\int_0^1 \dots \int_0^1 [A(1-z)(1-s_1)(1-s_2)\dots(1-s_{n-1}) + Bzs_1 s_2 \dots s_{n-1}]^{-(d-n+1)}$$
$$\times [s_1(1-s_1)]^{d-\frac{n+1}{2}}[s_2(1-s_2)]^{d-\frac{n+2}{2}}\dots[s_{n-1}(1-s_{n-1})]^{d-\frac{2n-1}{2}} ds_1 ds_2 \dots ds_{n-1} \qquad \text{......(15)},$$

where

$$H = K\prod_{i=0}^{n-1}\Gamma\left(\frac{2d-i}{2}\right)A^{\frac{M-n}{2}}B^{\frac{N-n}{2}}\frac{\psi^{\frac{M-n-2}{2}}}{(1+\psi)^{d-n+1}}, \quad d = \frac{M+N-2}{2}, \quad z = \frac{1}{1+\psi}.$$

Without loss of generality, we can assume that $B \leqslant A$. For, if $A < B$ we can make the transformation $s_i = 1 - \bar{s}_i$, $z = 1 - \bar{z}$, where $\bar{z} = \frac{\psi}{1+\psi}$, and get the desired form. If we denote $\frac{B}{A}$ by ϵ, then we can write

$$F(\psi) = \prod_{i=1}^{n}\left[\frac{\Gamma\left(\frac{M+N-1-i}{2}\right)}{\Gamma\left(\frac{M-i}{2}\right)\Gamma\left(\frac{N-i}{2}\right)}\right]\epsilon^{\frac{N-n}{2}}\psi^{\frac{M-n-2}{2}}(1+\psi)^{-\left(\frac{M+n}{2}-n\right)}$$
$$\times\int_0^1\int_0^1\dots\int_0^1\left\{1 - \frac{(1-s_1)(1-s_2)\dots(1-s_{n-1})}{1+\psi}\right.$$
$$\left. - \left[1-(1-s_1)(1-s_2)\dots(1-s_{n-1}) - \frac{\epsilon s_1 s_2 \dots s_{n-1}}{1+\psi}\right]\right\}^{-\left(\frac{M+N}{2}-n\right)}$$
$$\times[s_1(1-s_1)]^{\frac{M+N-n-3}{2}}[s_2(1-s_2)]^{\frac{M+N-n-4}{2}}\dots[s_{n-1}(1-s_{n-1})]^{\frac{M+N-2n-1}{2}} ds_1 ds_2 \dots ds_{n-1} \qquad \text{......(16)}.$$

From the expression in the brace in the integrand, we have

$$\frac{1-(1-s_1)(1-s_2)\dots(1-s_{n-1}) - \frac{\epsilon s_1 s_2 \dots s_{n-1}}{1+\psi}}{1 - \frac{(1-s_1)(1-s_2)\dots(1-s_{n-1})}{1+\psi}} < 1$$

and

$$\frac{(1-s_1)(1-s_2)\dots(1-s_{n-1})}{1+\psi} < 1$$

for $0 \leqslant s_i \leqslant 1$ and $\psi > 0$. Hence, the quantity in the brace can be expanded into a double infinite series which is uniformly convergent in the s's in the region of integration, and the integrations can be performed term by term.

For practical purposes, however, we can find the distribution of ψ for $n=2$ by means of (12 *b*). Indeed, the joint distribution of ξ and η for this case is

$$\frac{2^{M+N-6} A_2^{\frac{M-2}{2}} B_2^{\frac{N-2}{2}} \xi^{\frac{M-4}{2}} \eta^{\frac{N-4}{2}}}{\Gamma(M-2)\,\Gamma(N-2)} e^{-2\sqrt{A_2\xi}-2\sqrt{B_2\eta}}\, d\xi\, d\eta \quad\text{............(17).}$$

Setting $\frac{\xi}{\eta}=\psi$ and integrating with respect to η, we get

$$F(\psi)=\frac{\Gamma(M+N-4)\, A_2^{\frac{M-2}{2}} B_2^{\frac{N-2}{2}} \psi^{\frac{M-4}{2}}}{2\Gamma(M-2)\,\Gamma(N-2)\,(\sqrt{A_2\psi}+\sqrt{B_2})^{M+N-4}} \quad\text{............(18).}$$

If the two samples are from populations with $\alpha=\beta$, then $F(\psi)$ will only involve M, N and ψ. The condition $\alpha=\beta$, however, does not imply that the standard deviations and the correlations of the two variates for the two populations are identical.

The foregoing analysis can be extended to two samples drawn from populations with different numbers of variates. The moments of the ratio of the two generalized variances can be readily inferred from (13). To consider the simplest case of this kind, let ξ be the variance in a sample of one variate and η the generalized variance in a sample of two variates. It is reasonable to use as our statistical coefficient the ratio $\frac{\xi}{\sqrt{\eta}}=\theta$, instead of $\frac{\xi}{\eta}$. The distribution of θ is

$$f(\theta)=\frac{2^{N-2} A_1^{\frac{M-1}{2}} B_2^{\frac{N-2}{2}}\, \Gamma\left(\frac{M+2N-5}{2}\right) \theta^{\frac{M-3}{2}}}{\Gamma\left(\frac{M-1}{2}\right)\Gamma(N-2)\,(A_1\theta+2\sqrt{B_2})^{\frac{M+2N-5}{2}}} \quad\text{............(19).}$$

5. *Ratio of independent Generalized Variance to any of its principal Minors.* Here we shall consider the moments and distribution of the ratio of $|a_{ij}|$ in (7) to any one of its principal minors of tth order. Without loss of generality, we can take as our minor the one standing in the upper left corner of $|a_{ij}|$, because any principal minor can be shifted to that position by proper interchange of rows and columns in the determinant accompanied by similar changes in A to maintain the usual correspondence between the statistical coefficients and population parameters. Denote the ratio

$$\frac{|a_{ij}|}{|a_{pq}|}, \quad (i,j=1,2,\dots n;\ p,q=1,2,\dots t;\ t<n),$$

by ϕ. The kth moment of ϕ is

$$M_k(\phi)=\frac{A^{\frac{N-1}{2}}}{\pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{N-i}{2}\right)} \int e^{-\sum_{i,j=1}^{n} A_{ij}a_{ij}} |a_{ij}|^{\frac{N-n-2}{2}+k} |a_{pq}|^{-k}\, da \quad\text{...(20).}$$

We remark that the result of integrating (7) with respect to $a_{i,n}$ $(i=1,2,\dots n)$ is to reduce it to the distribution of the variances and covariances of the first $n-1$ variates of the n-variate sample. If we integrate further with respect to

$$a_{j,n-1}, \quad (j=1,2,\dots n-1),$$

we reduce it to the case of an $(n-2)$-variate population. If the process be performed $n-t$ times, the distribution is reduced to the case of a t-variate population. By argument similar to that used in deducing (10) from (9) we find

$$\int e^{-\sum_{i,j=1}^{n} A_{ij}a_{ij}} |a_{ij}|^{\frac{N-n-2}{2}+k} |a_{pq}|^{-k} da_{n-t}$$

$$= \frac{|A_{pq}^{(t)}|^{\frac{N-1}{2}+k}}{A^{\frac{N-1}{2}+k}} \pi^{\frac{n(n-1)}{4}-\frac{t(t-1)}{4}} \prod_{i=t+1}^{n} \Gamma\left(\frac{N-i}{2}+k\right) e^{-\sum_{p,q=1}^{t} A_{pq}^{(t)}a_{pq}} |a_{pq}|^{\frac{N-t-2}{2}} \quad \text{......(21)},$$

where the integration is performed with respect to all variables except those contained in $|a_{pq}|$, $A_{pq}^{(t)} = \dfrac{N\Delta_{pq}^{(t)}}{2\sigma_p\sigma_q\Delta^{(t)}}$ $(p, q = 1, 2, \dots t)$, $\Delta^{(t)}$ is the principal minor of order t in the upper left corner of Δ and $\Delta_{pq}^{(t)}$ is the co-factor corresponding to ρ_{pq} in $\Delta^{(t)}$. If (21) be integrated with respect to the variables a_{pq} $(p, q = 1, 2, \dots t)$ and the result multiplied by

$$\frac{|A_{ij}|^{\frac{N-1}{2}}}{\pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{N-i}{2}\right)}$$

we obtain

$$M_k(\phi) = B^{-k} \frac{\prod_{i=t+1}^{n} \Gamma\left(\frac{N-i}{2}+k\right)}{\prod_{i=t+1}^{n} \Gamma\left(\frac{N-i}{2}\right)},$$

where

$$B = \frac{N^{n-t}\Delta^{(t)}}{2^{n-t}\sigma^2_{t+1}\dots\sigma_n^2\Delta}.$$

Therefore, from (2) the distribution $f(\phi)$ is

$$f(\phi) = \frac{B^{\frac{N-n}{2}} \phi^{\frac{N-n}{2}-1}}{\prod_{i=t+1}^{n} \Gamma\left(\frac{N-i}{2}\right)} \int_0^\infty \int_0^\infty \dots \int_0^\infty (v_1 v_2 \dots v_{n-t-1})^{\frac{1}{2}}$$

$$\times e^{-v_1-\frac{v_2}{v_1}-\dots-\frac{v_{n-t-1}}{v_{n-t-2}}-\frac{B\phi}{v_{n-t-1}}} dv_1 dv_2 \dots dv_{n-t-1} \quad \text{............(22)}.$$

When $t = n-1$, we have the distribution of the variance of the difference between the nth variate and its estimate from the regression "plane" of the remaining $n-1$ variates, that is

$$f_1(\phi) = \frac{B_1^{\frac{N-n}{2}}}{\Gamma\left(\frac{N-n}{2}\right)} \phi^{\frac{N-n}{2}-1} e^{-B_1\phi} \quad \text{.....................(22 } a\text{)},$$

where $B_1 = \dfrac{N}{2\sigma_n^2(1-R^2)}$ and R is the multiple correlation coefficient between the nth variate and the first $n-1$ variates.

For $t=n-2$ $$f_2(\phi)=\frac{2^{N-n-1}B_2^{\frac{N-n}{2}}}{\Gamma(N-n)}\phi^{\frac{N-n}{2}-1}e^{-2\sqrt{B_2\phi}} \quad\text{..................(22b),}$$

where $$B_2=\frac{N^2\Delta^{(n-2)}}{4\sigma^2_{n-1}\sigma_n^2\Delta}.$$

6. *Generalization of the Correlation Ratio.* Let p samples ω_β ($\beta=1,2,\dots p$) of N_β items respectively be drawn from a normal population of one variable and let $\bar{x}_\beta$ and s_β^2 be the mean and variance of ω_β. Let Ω be the sample formed by pooling the ω's and let its mean and variance be denoted by $\bar{X}$ and S^2. The statistical coefficient η, defined as

$$\eta^2=\frac{\sum\limits_{\beta=1}^{p} N_\beta(\bar{x}_\beta-\bar{X})^2}{NS^2}, \qquad \left(\sum_{\beta=1}^{p}N_\beta=N\right) \quad\text{..................(23),}$$

is known as the correlation ratio with the samples ω_β forming the p categories of the independent variable. The distribution of η^2 defined in this manner was first found by Fisher* by his analysis of variance, and later by Hotelling† by a different method.

In this section we shall generalize the above definition of η^2 for samples from an n-variate normal population, and find its moments and distribution. We note from (23) that η^2 is the ratio of the weighted variance of the means of the p sub-samples ω_β to the variance of Ω. Now, let us suppose p samples ω_β' ($\beta=1,2,\dots p$) of n_β items respectively to be drawn from an n-variate normal population. Let the sample formed by pooling the ω''s be Ω', which will have $\sum\limits_{\beta=1}^{p} n_\beta=N$ items. The statistical coefficient to be considered is the ratio of the generalized weighted variance of the means of the ω''s to the generalized variance of Ω'. That is

$$U=\frac{|b_{ij}|}{|a_{ij}|},$$

where $$b_{ij}=b_{ji}=\frac{1}{N}\sum_{\beta=1}^{p}n_\beta(\bar{X}_{i\beta}-\bar{X}_i)(\bar{X}_{j\beta}-\bar{X}_j),$$

and $$a_{ij}=a_{ji}=\frac{1}{N}\sum_{\beta=1}^{p}\sum_{\alpha=1}^{n_\beta}(x_{i\beta\alpha}-\bar{X}_i)(x_{j\beta\alpha}-\bar{X}_j),$$

where $\bar{X}_{i\beta}$ is the mean of the ith variate in the βth sample and $x_{i\beta\alpha}$ is the value of the αth individual for the ith variate in the βth sample ω_β'. We observe that a_{ij} can be written as $b_{ij}+c_{ij}$, where

$$c_{ij}=\frac{1}{N}\sum_{\beta=1}^{p}\sum_{\alpha=1}^{n_\beta}(x_{i\beta\alpha}-\bar{X}_{i\beta})(x_{j\beta\alpha}-\bar{X}_{j\beta}) \quad\text{..................(24),}$$

or setting $$\sum_{\alpha=1}^{n_\beta}(x_{i\beta\alpha}-\bar{X}_{i\beta})(x_{j\beta\alpha}-\bar{X}_{j\beta})=n_\beta v_{ij\beta},$$

* R. A. Fisher: *Proceedings of the International Mathematical Congress*, Vol. II. Toronto (1924).

† H. Hotelling: "The Distribution of Correlation Ratios calculated from random Data," *Proceedings of the National Academy of Sciences*, Vol. XI. No. 10 (1925), pp. 657—662.

we have $$c_{ij}=\frac{1}{N}\sum_{\beta=1}^{p} n_\beta v_{ij\beta}, \qquad (i,j=1,2,\dots n) \quad \text{......(24 a)}.$$

Clearly $v_{ij\beta}$ is the general element in the determinant of variances and covariances in the sample ω_β'. The moments of $|a_{ij}|$ can be written down at once from (10) as

$$A^{-k}\prod_{i=1}^{n}\left[\frac{\Gamma\left(\frac{N-i}{2}+k\right)}{\Gamma\left(\frac{N-i}{2}\right)}\right].$$

It is well known that the system of means is distributed independently of the system of variances and covariances in samples from an n-variate normal population. Therefore, the system $\{b_{ij}\}$* is distributed independently of the set $\{c_{ij}\}$. The means in a sample of N individuals are distributed according to

$$\frac{A^{\frac{1}{2}}}{\pi^{\frac{n}{2}}} e^{-\sum_{i,j=1}^{n} A_{ij}(\bar{X}_i-m_i)(\bar{X}_j-m_j)}\, d\bar{X} \quad \text{..........................(25)},$$

where A_{ij} is given in (7) and m_i is the mean of x_i in the population, which, except for a factor N in each A_{ij}, is the distribution of the parent population. Therefore, the distribution of the set of statistical coefficients $\{b_{ij}\}$ can be deduced from (7), for $p>n$, as

$$\frac{\underline{p-1}}{\pi^{\frac{n(n-1)}{}}\prod_{i=1}^{n}\Gamma\left(\frac{p-i}{2}\right)}\cdot e^{-\sum_{i,j=1}^{n}A_{ij}b_{ij}}\,|b_{ij}|^{\frac{p-n-2}{2}}\,db \quad \text{..............(26)}.$$

The distribution of the variances and covariances $\{v_{ij\beta}\}$ in (24 a) is given by (7) with N replaced by n_β. Hence, it can be shown without much difficulty that the distribution of the set $\{c_{ij}\}$ is given by

$$\frac{A^{\frac{N-p}{2}}}{\pi^{\frac{n(n-1)}{4}}\prod_{i=1}^{n}\Gamma\left(\frac{N-p+1-i}{2}\right)}\, e^{-\sum_{i,j=1}^{n}A_{ij}c_{ij}}\,|c_{ij}|^{\frac{N-p-n-1}{2}}\,dc \quad \text{.........(27)}.$$

One way to prove (27) is to evaluate the characteristic function of the set $\{c_{ij}\}$ from the distributions of the quantities $\{v_{ij\beta}\}$, $(\beta=1,2,\dots p)$. Indeed, we define the characteristic function $\phi(\bar{\alpha})$ of $\{c_{ij}\}$ as

$$\phi(\bar{\alpha})=\frac{\prod_{\beta=1}^{p}|A_{ij}^{(\beta)}|^{\frac{n_\beta-1}{2}}}{\pi^{\frac{pn(n-1)}{2}}\prod_{\beta=1}^{p}\prod_{i=1}^{n}\Gamma\left(\frac{n_\beta-i}{2}\right)}\int e^{-\sum_{\beta=1}^{p}\sum_{i,j=1}^{n}\left(A_{ij}^{(\beta)}-\frac{\alpha_{ij}n_\beta}{N(2-\delta_{ij})}\right)v_{ij\beta}}$$

$$\times\prod_{\beta=1}^{p}|v_{ij\beta}|^{\frac{n_\beta-n-2}{2}}\,dV \quad \text{...(28)},$$

where dV is the product of the differentials of all the v's, $\bar{\alpha}$ is the set $\{\alpha_{ij}\}$, δ_{ij} is the

* The brace notation, $\{\ \}$, will be used to indicate a system or aggregate, as distinct from the notation $|\ \ |$ used to indicate a determinant.

Kronecker delta which is unity for $i=j$ and zero for $i \neq j$, $A_{ij}^{(\beta)} = \frac{n_\beta}{N} A_{ij}$ and $\alpha_{ij} \equiv \alpha_{ji}$. This integral breaks up into p integrals, each of the form (8). Applying the results of (8) on each of the integrals, we get

$$\phi(\bar{\alpha}) = |A_{ij}|^{\frac{N-p}{2}} \left| A_{ij} - \frac{\alpha_{ij}}{2-\delta_{ij}} \right|^{-\frac{N-p}{2}} \quad \text{.(29).}$$

This is clearly the characteristic function belonging to (27). Hence (27) is the distribution of the system $\{c_{ij}\}$. We are now in a position to state that

$$\int e^{-\sum\limits_{i,j=1} A_{ij}(b_{ij}+c_{ij})} |a_{ij}|^k |b_{ij}|^{\frac{p-n-2}{2}} |c_{ij}|^{\frac{N-p-n-1}{2}} \, db\, dc = \frac{A^{-k}}{H} \prod_{i=1}^{n} \frac{\Gamma\left(\frac{N-i}{2}+k\right)}{\Gamma\left(\frac{N-i}{2}\right)} \quad \text{......(30),}$$

where
$$H = \frac{A^{\frac{N-1}{2}}}{\pi^{\frac{n(n-1)}{2}} \prod\limits_{i=1}^{n} \left[\Gamma\left(\frac{p-i}{2}\right)\Gamma\left(\frac{N-p+1-i}{2}\right)\right]},$$

which is the product of the constant coefficients of the distributions of the two sets of statistical coefficients $\{b_{ij}\}$ and $\{c_{ij}\}$. If in (30) we set $k=-h$, $p=p+2h$ and $N=N+2h$, afterwards multiplying by H, we have for the hth moment of U

$$M_h(U) = \prod_{i=1} \frac{\Gamma\left(\frac{N-i}{2}\right)\Gamma\left(\frac{p-i}{2}+h\right)}{\Gamma\left(\frac{p-i}{2}\right)\Gamma\left(\frac{N-i}{2}+h\right)}$$

The distribution $\phi(U)$ of U satisfies an integral equation of type (B) and from (5) we have

$$\phi(U) = \frac{\prod\limits_{i=1}^{n} \Gamma\left(\frac{N-i}{2}\right) U^{\frac{p-n-2}{2}} (1-U)^{\frac{n(N-p)}{2}-1}}{\prod\limits_{i=1}^{n} \left[\Gamma\left(\frac{p-i}{2}\right)\right] \Gamma^n\left(\frac{N-p}{2}\right)} \int_0^1 \int_0^1 \cdots \int_0^1 (v_1 v_2 \ldots v_{n-1})^{\frac{N-p}{2}-1}$$

$$\times (1-v_1)^{\frac{(n-1)(N-p)}{2}-1} (1-v_2)^{\frac{(n-2)(N-p)}{2}-1} \ldots (1-v_{n-1})^{\frac{N-p}{2}-} \quad \ldots)]^{-\left(\frac{N-p-1}{2}\right)}$$

$$\times [1-\{v_1+v_2.(1-v_1)\}(1-U)]^{-\left(\frac{N-p-1}{2}\right)} \ldots$$

$$\times [1-\{v_1+v_2.(1-v_1)+\ldots+v_{n-1}.(1-v_1)\ldots(1-v_{n-2})\}(1-U)]^{-\left(\frac{N-p-1}{2}\right)}$$

$$\times dv_1 dv_2 \ldots dv_{n-1} \quad \text{...................(31).}$$

The range of U is from 0 to 1, since $B=1$ in (5) for this case.

For $n=1$ we get

$$\phi_1(U) = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{N-p}{2}\right)} U^{\frac{p-3}{2}} (1-U)^{\frac{N-p-2}{2}} \quad \text{..(31a),}$$

the well-known [redacted]tained by Fish[redacted] and Hotelling [redacted]ution of η^2

For $n=2$

$$\phi_2(U)=\frac{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-2}{2}\right)U^{\frac{p-4}{2}}(1-U)^{N-p-1}}{\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{p-2}{2}\right)\Gamma^2\left(\frac{N-p}{2}\right)}$$

$$\times\int_0^1[v_1(1-v_1)]^{\frac{N-p}{2}-1}[1-v_1(1-U)]^{-\left(\frac{N-p-1}{2}\right)}dv_1$$

$$=\frac{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-2}{2}\right)U^{\frac{p-4}{2}}(1-U)^{N-p-1}}{\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{p-2}{2}\right)\Gamma(N-p)}F\left(\frac{N-p-1}{2},\frac{N-p}{2},N-p,1-U\right) \quad \text{......(31 b).}$$

7. *Generalization of* $1-\eta^2$. In the case of samples of one variable the distribution of $1-\eta^2$ can be found from that of η^2 by a simple change of variable, but such is not true in the generalization which we shall consider.

From (23) we find that $$1-\eta^2=\frac{\sum_{\beta=1}^{p} n_\beta s_\beta^2}{NS^2} \quad \text{..............................(32),}$$

that is, the ratio of the weighted mean of the variances of the samples ω_β to the variance of Ω. The quantity which we shall consider as a generalization of $1-\eta^2$ is

$$W=\frac{|c_{ij}|}{|a_{ij}|} \quad \text{..................................(33),}$$

$\{c_{ij}\}$ and $\{a_{ij}\}$ have been defined in Section 6.

be shown that W arises as the maximum likelihood criterion λ_H of the type used by Neyman and E. S. Pearson for testing an hypothesis H that p samples ω_β are from a subclass d of a class D of admissible populations. In the present case D is the class of all sets of p n-variate normal populations in each set of which the corresponding variances and covariances are the same, but the means are completely independent, while d is the subclass in each set of which the means are the same, that is to say in each set of which the p populations are identical. The maximum of the L-function of the samples ω_β ($\beta=1, 2, \dots p$) for the populations of class D is

$$M_D=J\,|c_{ij}|^{-\frac{N}{2}}\prod_{\beta=1}^{p}|v_{ij\beta}|^{\frac{n_\beta-n-2}{2}}.$$

The maximum of the L-function of the samples from populations of class d is

$$M_d=J\,|a_{ij}|^{-\frac{N}{2}}\prod_{\beta=1}^{p}|v_{ij\beta}|^{\frac{n_\beta-n-2}{2}},$$

where J is a constant depending on n_β ($\beta=1, 2, \dots p$) and n. Therefore

$$\lambda_H=\frac{M_d}{M_D}=W^{\frac{N}{2}}.$$

The hth moment of W, when H is true, is found from (30) by setting $k=-h$, $N=N+2h$, and then multiplying by

$$\frac{A^{\frac{N-1}{2}}}{\pi^{\frac{n(n-1)}{2}}\prod_{i=1}^{n}\left[\Gamma\left(\frac{p-i}{2}\right)\Gamma\left(\frac{N-p+1-i}{2}\right)\right]}.$$

Accordingly, we get

$$M_h(W)=\prod_{i=1}\left[\frac{\Gamma\left(\frac{N-i}{2}\right)\Gamma\left(\frac{N-p+1-i}{2}+h\right)}{\Gamma\left(\frac{N-i}{2}+h\right)\Gamma\left(\frac{N-p+1-i}{2}\right)}\right] \quad .(34).$$

The distribution of W is clearly of the form (5) and can be written as

$$\theta(W)=\frac{\prod_{i=1}^{n}\Gamma\left(\frac{N-i}{2}\right)W^{\frac{N-p-n+1}{2}-1}(1-W)^{\frac{n(p-1)}{2}-1}}{\prod_{i=1}^{n}\Gamma\left(\frac{N-p+1-i}{2}\right)\Gamma^{n}\left(\frac{p-1}{2}\right)}\int_0^1\int_0^1\dots\int_0^1(v_1v_2\dots v_{n-1})^{\frac{p-2}{2}}$$

$$\times(1-v_1)^{\frac{(n-1)(p-1)}{2}-1}(1-v_2)^{\frac{(n-2)(p-1)}{2}-1}\dots(1-v_{n-1})^{\frac{p-1}{2}-1}[1-v_1(1-W)]^{-\frac{p-2}{2}}$$

$$\times[1-\{v_1+v_2.(1-v_1)\}(1-W)]^{-\frac{p-2}{2}}\dots$$

$$\times[1-\{v_1+v_2.(1-v_1)+\dots+v_{n-1}.(1-v_1).(1-v_2)\dots(1-v_{n-2})\}(1-W)]^{-\frac{p-2}{2}}$$

$$\times dv_1dv_2\dots dv_{n-1} \quad \dots\dots(35).$$

The range of W is from 0 to 1.

For $n=1$ we find

$$\theta_1(W)=\frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-p}{2}\right)\Gamma\left(\frac{p-1}{2}\right)}W^{\frac{N-p-2}{2}}(1-W)^{\frac{p-3}{2}} \quad \dots\dots\dots(35\,a),$$

the distribution of $1-\eta^2$.

For $n=2$

$$\theta_2(W)=\frac{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-2}{2}\right)}{\Gamma\left(\frac{N-p}{2}\right)\Gamma\left(\frac{N-p-1}{2}\right)\Gamma(p-1)}W^{\frac{N-p-3}{2}}(1-W)^{p-2}$$

$$\times F\left[\frac{p-2}{2},\frac{p-1}{2},p-1,1-W\right] \quad \dots\dots(35\,b).$$

At this point we remark that since the elements $\{b_{ij}\}$ are distributed independently of the quantities $\{c_{ij}\}$ and since the distribution of each system is essentially of the form (7), we can deduce from (13) that the kth moment of the ratio $\frac{b_{ij}}{c_{ij}}$ is

$$\prod_{i=1}\frac{\Gamma\left(\frac{p-i}{2}+k\right)\Gamma\left(\frac{N-p+1-i}{2}-k\right)}{\Gamma\left(\frac{p-i}{2}\right)\Gamma\left(\frac{N-p+1-i}{2}\right)}$$

and its distribution can be found from (16).

8. *Generalization of "Student's" Ratio.* For samples of a single variable "Student's" ratio is defined as the quantity

$$z = \frac{(\bar{x} - m)}{s},$$

where s is the standard deviation, $\bar{x}$ the mean of the sample and m the population mean. In a recent paper, Hotelling* has generalized the statistical coefficient z^2 for samples from an n-variate normal population and has found the distribution of T^2 which is the product of the generalized z^2 and the number of degrees of freedom in the sample. However, we shall show that the distribution of this generalized ratio can be reached by the methods used in Sections 6 and 7.

In a sample of N individuals from an n-variate normal population, let the set of variances and covariances $\{a_{ij}\}$ be defined as in Section 3. Let the sample means be $\{\bar{X}_i\}$ and the corresponding population means be $\{m_i\}$. The distribution of the set $\{a_{ij}\}$ is given by (7) and that of $\{\bar{X}_i\}$ by (25). The statistical coefficient which we shall consider first is

$$\frac{|a_{ij}|}{|e_{ij}|} \qquad (36),$$

where
$$e_{ij} = a_{ij} + (\bar{X}_i - m_i)(\bar{X}_j - m_j).$$

It is not difficult to show that Y arises as the maximum likelihood criterion $\lambda_{H'}$ for testing the hypothesis H' that our sample is from a subset d' of n-variate populations D', where d' is the class of normal populations having a specified set of means $\{m_i\}$ and any set of variances and covariances, and D' is the class of all n-variate normal populations with any set of means, variances and covariances. Proceeding as in Section 6 for λ_H, we find

$$\lambda_{H'} = \left[\frac{|a_{ij}|}{|e_{ij}|}\right]^{\frac{N}{2}}.$$

By the procedure used in finding the distribution of $\{c_{ij}\}$ in (27), we can show that the distribution of the set $\{e_{ij}\}$ is

$$\frac{A^{\frac{N}{2}}}{\pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{N+1-i}{2}\right)} e^{-\sum_{i,j=1}^{n} A_{ij} e_{ij}} |e_{ij}|^{\frac{N-n-1}{2}} de \quad \ldots\ldots\ldots\ldots(37).$$

Therefore, the kth moment of $|e_{ij}|$ is the same as that of ξ in (10) with N replaced by $N+1$. Since the means $\{\bar{X}_i\}$ and the system $\{a_{ij}\}$ are independently distributed, we shall have

$$\int e^{-\sum_{i,j=1}^{n} A_{ij}[a_{ij} + (\bar{X}_i - m_i)(\bar{X}_j - m_j)]} |e_{ij}|^k |a_{ij}|^{\frac{N-n-2}{2}} da\, d\bar{X}$$

$$= A^{-k-\frac{N}{2}} \pi^{\frac{n(n-1)}{4} + \frac{n}{2}} \prod_{i=1}^{n} \left[\frac{\Gamma\left(\frac{N-i}{2}\right) \Gamma\left(\frac{N+1-i}{2} + k\right)}{\Gamma\left(\frac{N+1-i}{2}\right)}\right] \ldots\ldots\ldots\ldots(38).$$

* H. Hotelling: *Annals of Mathematical Statistics*, Vol. II. (1931).

Changing N to $N+2h$ and k to $-h$ and multiplying by

$$\frac{A^{\frac{N}{2}}}{\pi^{\frac{n(n-1)}{4}+\frac{n}{2}} \prod_{i=1}^{n} \Gamma\left(\frac{N-i}{2}\right)}$$

we get

$$M_h(Y) = \frac{\Gamma\left(\frac{N}{2}\right)\Gamma\left(\frac{N-n}{2}+h\right)}{\Gamma\left(\frac{N}{2}+h\right)\Gamma\left(\frac{N-n}{2}\right)} \quad \text{.......................(39).}$$

Hence, from (5), the distribution of Y is

$$F(Y)\,dY = \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-n}{2}\right)\Gamma\left(\frac{n}{2}\right)} Y^{\frac{N-n}{2}-1}(1-Y)^{\frac{n}{2}-1}\,dY \text{...........(40),}$$

with a range from 0 to 1.

By breaking up the rows or columns of $|e_{ij}|$ and expressing $|e_{ij}|$ as a sum of the resulting determinants, it can be readily shown that

$$Y = \frac{1}{1+\frac{T^2}{N-1}},$$

where

$$\frac{T^2}{N-1} = \sum_{p,q=1}^{n} \frac{\bar{a}_{pq}}{|a_{ij}|}(\bar{X}_p - m_p)(\bar{X}_q - m_q),$$

and $\bar{a}_{pq}$ is the co-factor of a_{pq} in $|a_{ij}|$.

Making the change of variable from Y to T in (40), we find the distribution of T to be

$$\frac{2\Gamma\left(\frac{N}{2}\right)}{\left(\frac{N-n}{2}\right)\Gamma\left(\frac{n}{2}\right)(N-1)^{\frac{n}{2}}} \frac{T^{n-1}\,dT}{\left(1+\frac{T^2}{N-1}\right)^{\frac{N}{2}}} \quad \text{.(41),}$$

which is the distribution established by Hotelling.

We note that (41) has been derived without making use of the property of the invariance of T^2 under all homogeneous linear transformations of the n variates in the population. This property, however, played an important part in Hotelling's derivation.

9. *Generalization of the λ-Criterion appropriate to k Samples.* In 1931, E. S. Pearson and Neyman* considered a certain maximum likelihood criterion λ_H for testing the hypothesis H that k samples are drawn from a subclass ω of a class Ω of admissible populations, where Ω is the class of all sets of k univariate normal

* J. Neyman and E. S. Pearson: *Bulletin de l'Académie Polonaise des Sciences et des Lettres, Série A, Sciences mathématiques*, 1930 and 1931.

$\lambda_{H(n)}$ and $\lambda_{H'(n)}$ of this Section and the λ_H of Section 7 are respectively generalizations for the case of n variables of the λ_H, λ_{H_1} and λ_{H_2} used by these writers in the case of a single variable.

populations and ω is the subclass in each set of which the k populations have the same means and standard deviations. This criterion, for the case of k samples of one variable, was found in the following manner:

The maximum of the L-function of all k samples from populations of class Ω is

$$M_\Omega = C(S_0^2)^{-\frac{N}{2}} \prod_{\beta=1}^{k} (s_\beta^2)^{\frac{n_\beta-3}{2}},$$

where C is a constant depending on the n_β's ($\beta = 1, 2, \ldots k$), the numbers of individuals in the samples, and $N = \sum_{\beta=1}^{k} n_\beta$, s_β^2 is the variance of the βth sample and S_0^2 the variance of the pool of the k samples.

The maximum of the joint likelihood of the k samples from populations of class ω is

$$M_\omega = C \prod_{\beta=1}^{k} (s_\beta^2)^{-\frac{n_\beta}{2}} \prod_{\beta=1}^{k} (s_\beta^2)^{\frac{n_\beta-3}{2}}.$$

For the ratio of these maximums we have

$$\lambda_H = \frac{M_\omega}{M_\Omega} = \prod_{\beta=1}^{k} \left(\frac{s_\beta^2}{S_0^2}\right)^{\frac{n_\beta}{2}},$$

which is the criterion adopted by E. S. Pearson and Neyman for testing H.

The generalization of this criterion for testing hypothesis H on k samples from an n-variate normal population is straightforward. Indeed, the generalized criterion is

$$\lambda_{H(n)} = \prod_{\beta=1}^{k} \left[\frac{|s_{ij\beta}|}{|S_{ij0}|}\right]^{\frac{n_\beta}{2}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(42),$$

where $|s_{ij\beta}|$ is the generalized variance of the βth sample and $|S_{ij0}|$ is the generalized variance of the sample formed by combining the k samples.

To find the moments of $\lambda_{H(n)}$, when the hypothesis is, we proceed as in Sections 6 and 7, and deduce at once from (10) the tth moment of $|S_{ij0}|^{\frac{1}{2}N}$ $\left(N = \sum_{\beta=1}^{k} n_\beta\right)$ to be

$$A^{-\frac{Nt}{2}} \prod_{i=1}^{n} \left[\frac{\Gamma\left(\frac{N(1+t)-i}{2}\right)}{\Gamma\left(\frac{N-i}{2}\right)}\right] \qquad .(43).$$

But, since the elements $\{S_{ij0}\}$ are expressible in terms of the variances, covariances and means of the k samples under consideration, we must have

$$\int e^{-\sum_{i,j=1}^{n}\sum_{\beta=1}^{k} A_{ij}^{(\beta)}[s_{ij\beta}+(\bar{X}_{i\beta}-m_i)(\bar{X}_{j\beta}-m_j)]} \prod_{\beta=1}^{k} |s_{ij\beta}|^{\frac{n_\beta-n-2}{2}} |S_{ij0}|^{\frac{Nt}{2}}\, ds\, d\bar{X}$$

$$A^{-\frac{Nt}{2}} \prod_{\beta=1}^{k} A_\beta^{-\frac{n_\beta}{2}} \pi^{\frac{kn(n-1)}{4}+\frac{kn}{2}} \prod_{i=1}^{n} \left[\frac{\prod_{\beta=1}^{k} \Gamma\left(\frac{n_\beta-i}{2}\right) \Gamma\left(\frac{N(1+t)-i}{2}\right)}{\Gamma\left(\frac{N-i}{2}\right)}\right]$$

.(44),

where A_β is the determinant A with N replaced by n_β and $A_{ij(\beta)} = \frac{n_\beta}{N} A_{ij}$. Setting $n_\beta = n_\beta(1+h)$, $N = N(1+h)$ and $t = -\frac{h}{1+h}$, and afterwards multiplying (44) by

$$\prod_{\beta=1}^{k}\left[\frac{A_\beta^{\frac{n_\beta}{2}}}{\pi^{\frac{n(n-1)}{4}+\frac{n}{2}}\prod_{i=1}^{n}\Gamma\left(\frac{n_\beta-i}{2}\right)}\right],$$

we get for the hth moment of $\lambda_{H(n)}$

$$M_h(\lambda_{H(n)}) = \prod_{\beta=1}^{k}\left[\left(\frac{N}{n_\beta}\right)^{\frac{nhn_\beta}{2}}\prod_{i=1}^{n}\left(\frac{\Gamma\left(\frac{n_\beta(1+h)-i}{2}\right)}{\Gamma\left(\frac{n_\beta-i}{2}\right)}\right)\right]\prod_{i=1}^{n}\left[\frac{\Gamma\left(\frac{N-i}{2}\right)}{\Gamma\left(\frac{N(1+h)-i}{2}\right)}\right] \quad \text{.........(45).}$$

For the case $n=1$, we have

$$M_h(\lambda_{H(1)}) = N^{\frac{hN}{2}}\prod_{\beta=1}^{k}\left[\frac{\Gamma\left(\frac{n_\beta(1+h)-1}{2}\right)}{n_\beta^{\frac{1}{2}hn_\beta}\,\Gamma\left(\frac{n_\beta-1}{2}\right)}\right]\frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N(1+h)-1}{2}\right)},$$

which was found by E. S. Pearson and Neyman by direct integration.

Let us modify the hypothesis which yielded the criterion $\lambda_{H(n)}$ in (42), and suppose that $\bar{\omega}$ is the class of all sets of n-variate normal populations in each set of which the corresponding variances and covariances are the same, but the means are completely independent. Let the definition of Ω be unchanged. Then clearly $\bar{\omega}$ is a subclass of Ω. Let the hypothesis that the k samples are from the subclass ω of populations Ω be $H'_{(n)}$. Then the E. S. Pearson and Neyman criterion $\lambda_{H'(n)}$ appropriate to $H'_{(n)}$ is readily found to be

$$\lambda_{H'(n)} = \frac{M_{\bar{\omega}}}{M_\Omega} = \prod_{\beta=1}^{k}\left[\frac{|s_{ij\beta}|}{|c_{ij}|}\right]^{\frac{n_\beta}{2}},$$

where
$$c_{ij} = \frac{1}{N}\sum_{\beta=1}^{k} n_\beta s_{ij\beta}.$$

The distribution of the set $\{c_{ij}\}$ in this case is given by (27) with p replaced by k. Proceeding exactly as in the case of $\lambda_{H(n)}$, we find that the hth moment of $\lambda_{H'(n)}$ is

$$M_h(\lambda_{H'(n)}) = \prod_{\beta=1}^{k}\left[\left(\frac{N}{n_\beta}\right)^{\frac{nhn_\beta}{2}}\prod_{i=1}^{n}\left(\frac{\Gamma\left(\frac{n_\beta(1+h)-i}{2}\right)}{\Gamma\left(\frac{n_\beta-i}{2}\right)}\right)\right]\prod_{i=1}^{n}\left[\frac{\Gamma\left(\frac{N-k+1-i}{2}\right)}{\Gamma\left(\frac{N(1+h)-k+1-i}{2}\right)}\right] \quad \text{......(45 \textit{bis}).}$$

For the case $n=1$, we have

$$M_h(\lambda_{H'(1)}) = \prod_{\beta=1}^{k}\left(\frac{N}{n_\beta}\right)^{\frac{hn_\beta}{2}}\frac{\Gamma\left(\frac{n_\beta(1+h)-1}{2}\right)}{\Gamma\left(\frac{n_\beta-1}{2}\right)}\;\frac{\Gamma\left(\frac{N-k}{2}\right)}{\Gamma\left(\frac{N(1+h)-k}{2}\right)},$$

as found by E. S. Pearson and Neyman

10. *Moments and Distributions of Ratios of Determinants of Correlation Coefficients.* The distribution of the correlation coefficients r in samples from a normal population in which the correlation is zero was first suggested by "Student"* and later verified by Fisher†. From this distribution one can readily find that of $1-r^2$ which can be written as the determinant $\begin{vmatrix} 1 & r \\ r & 1 \end{vmatrix}$. In this section we shall first consider the moments and distribution of the determinant of the correlation coefficients in a sample from a normal population of n independent variables. That is, we shall find the moments and distribution of $|r_{ij}|$ ‡, where $r_{ij}=r_{ji}$ and $r_{ii}=1$. The distribution of variances and covariances in a sample from such a population is given by (7), where the ρ's are all made zero.

Hence, we have, corresponding to (8),

$$\int e^{-\sum_{i=1}^{n} A_i a_{ii}} |a_{ij}|^{\frac{N-n-2}{2}} da = \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \left[\Gamma\left(\frac{N-i}{2}\right) A_i^{-\frac{N-1}{2}} \right] \quad \text{......(46)},$$

where

$$A_i = \frac{N}{2\sigma_i^2}.$$

If we change the variables by the transformation

$$a_{ij} = r_{ij}\sqrt{a_{ii}a_{jj}}, \qquad (i, j = 1, 2, \dots n;\ i \neq j)$$

then

$$\int e^{-\sum_{i=1}^{n} A_i a_{ii}} (a_{11}a_{22}\dots a_{nn})^{\frac{N-3}{2}} |r_{ij}|^{\frac{N-n-2}{2}} da\,dr = \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \left[\Gamma\left(\frac{N-i}{2}\right) A_i^{-\frac{N-1}{2}} \right] \quad \text{......(47)}.$$

That the set $\{r_{ij}\}$ is distributed independently of the set $\{a_{ii}\}$ can be shown by evaluating the characteristic function of $\{a_{ii}\}$ which is known to be

$$\phi(\bar{\alpha}) = \prod_{i=1}^{n} \left[A_i^{\frac{N-1}{2}} (A_i - \alpha_i)^{-\frac{N-1}{2}} \right],$$

since the a's are variances in samples from independent populations. This characteristic function must also satisfy the relation

$$\phi(\bar{\alpha}) = \frac{1}{H} \int e^{-\sum_{i=1}^{n} (A_i - \alpha_i) a_{ii}} (a_{11}a_{22}\dots a_{nn})^{\frac{N-3}{2}} |r_{ij}|^{\frac{N-n-2}{2}} da\,dr,$$

where H is the quantity on the right side of (47). From Stekloff's theory it follows that

$$\int |r_{ij}|^{\frac{N-n-2}{2}} dr$$

* "Student": "The Probable Error of a Correlation Coefficient," *Biometrika*, Vol. VI. (1908—1909), pp. 302—310.

† R. A. Fisher: *Biometrika*, Vol. X. (1915).

‡ See Ragnar Frisch: "Correlation and Scatter in statistical Variables," *Nordisk Statistisk Tidskrift*, Vol. VIII. (1928), pp. 36—102.

In this paper Frisch considers the significance of the determinant $|r_{ij}|$ and its principal minors in the procedure of fitting linear regression equations to scatter diagrams with special reference to the matter of detecting the presence of irrelevant or uncorrelated variables. He refers to the quantity $+\sqrt{|r_{ij}|}$ as the *collective scatter coefficient* of the sample, but he does not consider the problem of finding its sampling moments and distribution.

is necessarily a constant independent of the a's. Thus, we can deduce from (47) that the kth moment of $|r_{ij}| = \omega$ is

$$M_k(\omega) = \frac{\Gamma^{n-1}\left(\frac{N-1}{2}\right) \prod_{i=2}^{n} \Gamma\left(\frac{N-i}{2}+k\right)}{\Gamma^{n-1}\left(\frac{N-1}{2}+k\right) \prod_{i=2}^{n} \Gamma\left(\frac{N-i}{2}\right)} \quad \text{.................(48).}$$

The distribution of ω can therefore be written from (5) as

$$f(\omega) = \frac{\Gamma^{n-1}\left(\frac{N-1}{2}\right) \omega^{\frac{N-n}{2}-1} (1-\omega)^{\frac{n(n-1)}{4}-1}}{\prod_{i=2}^{n} \left[\Gamma\left(\frac{N-i}{2}\right) \Gamma\left(\frac{i-1}{2}\right)\right]} \int_0^1 \int_0^1 \dots \int_0^1 v_1^{\frac{1}{2}-1} v_2^{1-1} \dots v_{n-2}^{\frac{n-2}{2}-1}$$

$$\times (1-v_1)^{\frac{n(n-1)}{4}-\frac{3}{2}} (1-v_2)^{\frac{n(n-1)}{4}-\frac{5}{2}} \dots (1-v_{n-2})^{\frac{n-3}{2}} \{1 - v_1(1-\omega)\}^{-\frac{1}{2}}$$

$$\times [1 - \{v_1 + v_2(1-v_1)\}(1-\omega)]^{-1} \dots$$

$$\times [1 - \{v_1 + v_2(1-v_1) + \dots + v_{n-2}(1-v_1)\dots(1-v_{n-3})\}(1-\omega)]^{-\frac{n-2}{2}}$$

$$\times dv_1 dv_2 \dots dv_{n-2} \quad \text{...(49).}$$

For $n = 2$ we have

$$f_2(\omega) = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{N-2}{2}\right)} \omega^{\frac{N-4}{2}} (1-\omega)^{-\frac{1}{2}} \quad \text{..............(49a),}$$

the well-known distribution of $1 - r^2$.

For $n = 3$

$$f_3(\omega) = \frac{\Gamma^2\left(\frac{N-1}{2}\right) \omega^{\frac{N-5}{2}} (1-\omega)^{\frac{1}{2}}}{\Gamma\left(\frac{3}{2}\right) \Gamma\left(\frac{N-2}{2}\right) \Gamma\left(\frac{N-3}{2}\right)} F\left[\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{3}{2}, 1-\omega\right] \quad \text{......(49b).}$$

At this point we shall introduce a slightly more general function of the r's and consider the ratio of ω to the product of k of its principal minors which are mutually exclusive. Without loss of generality, we can consider the k minors as placed corner to corner down the main diagonal of $|r_{ij}|$, such that each element in this diagonal (all equal to unity) is included in one of the minors.

Thus, let
$$Z = \frac{\omega}{\prod_{\beta=1}^{k} \omega_\beta} \quad \text{.....................................(50),}$$

where ω_β is the βth principal minor from the upper left corner of ω and contains the inter-correlations of p_β variates and $\sum_{\beta=1}^{k} p_\beta = n$. If we multiply and divide the quantity on the right in (50) by $\prod_{i=1}^{n} a_{ii}$, we have Z expressed as the ratio of the

generalized variance to the product of k of its principal minors, all of which are mutually exclusive and are such that every element of the main diagonal of $|a_{ij}|$ is contained in one of them. If (46) be integrated with respect to all of the a's not contained in this set of principal minors, which will be denoted collectively by $\bar{a}$, we must have

$$\int e^{-\sum_{i=1}^{n} A_i a_{ii}} |a_{ij}|^{\frac{N-n-2}{2}} d\bar{a}$$

$$= \frac{\pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{N-i}{2}\right)}{\prod_{\beta=1}^{k} \left[\pi^{\frac{p_\beta(p_\beta-1)}{4}} \prod_{\alpha=1}^{p_\beta} \Gamma\left(\frac{N-\alpha}{2}\right)\right]} e^{-\sum_{i=1}^{n} A_i a_{ii}} \prod_{\beta=1}^{k} |a_{ij\beta}|^{\frac{N-n_\beta-2}{2}} d(a-\bar{a}) \quad \text{......(51)},$$

where $|a_{ij\beta}|$ is the βth minor in $|a_{ij}|$, and $a-\bar{a}$ is the set of a's in (46) not contained in $\bar{a}$.

If both sides of (51) be multiplied by $\prod_{\beta=1}^{k} |a_{ij\beta}|^{-h}$, which is constant as far as the set $\bar{a}$ is concerned, and if N be replaced by $N+2h$, afterwards integrating with respect to $a-\bar{a}$, then multiplying by

$$\frac{(A_1 A_2 \dots A_n)^{\frac{N-1}{2}}}{\pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{N-1}{2}\right)},$$

we obtain the expression for the hth moment of Z, which is

$$M_h(Z) = \prod_{\beta=1}^{k} \prod_{\alpha=1}^{p_\beta} \left[\frac{\Gamma\left(\frac{N-\alpha}{2}\right)}{\Gamma\left(\frac{N-\alpha}{2}+h\right)}\right] \prod_{i=1}^{n} \left[\frac{\Gamma\left(\frac{N-i}{2}+h\right)}{\Gamma\left(\frac{N-i}{2}\right)}\right] \quad \text{......(52)}.$$

This is clearly the hth moment of a function satisfying an integral equation of type (B), and hence the distribution of Z is of the form (5), where Z ranges from 0 to 1. An important case of (52) arises when $\beta=2$, $p_1=n-1$, and $p_2=1$ which yields the hth moment of $1-R^2$, where R is the multiple correlation coefficient in a sample from a population with its multiple correlation coefficient zero. That is

$$M_h(Z_1) = \frac{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-n}{2}+h\right)}{\Gamma\left(\frac{N-1}{2}+h\right)\Gamma\left(\frac{N-n}{2}\right)}.$$

From this we get the result

$$f_1(Z_1) = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-n}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} Z_1^{\frac{N-n-2}{2}} (1-Z_1)^{\frac{n-1}{2}-1},$$

from which we can deduce the distribution of R^2, by the simple change of variable $Z_1=1-R^2$, a result originally obtained by Fisher.

The distribution of Z can be found for a slightly more general problem than the one we have just considered. Thus, suppose a sample of N items is drawn from each of k independent normal populations, where the βth population has p_β inter-correlated variates. Let v_β be the generalized variance of the βth sample and V the generalized variance of the k samples treated as one sample with

$$\sum_{\beta=1}^{k} p_\beta = n$$

variates. Then it can be shown in a straightforward manner by the foregoing method that the hth moment of

$$\frac{V}{\prod\limits_{\beta=1}^{k} v_\beta} = \bar{Z}$$

is identical with $M_h(Z)$.

The use of $\bar{Z}$ as a criterion for testing the hypothesis that the k samples are from k independent normal populations can be interpreted as an extension of the use of $1 - R^2$ as a criterion for testing the hypothesis that a sample of N items of a single variable and a sample of N items of $n-1$ variables are from independent populations. For example, the criterion $\bar{Z}_2$ appropriate to testing the hypothesis that a sample of N items of two variables and one of N items of $n-2$ variables are from independent populations, will have its hth moment given by (52) for the special case $\beta = 2$, $p_1 = 2$, and $p_2 = n-2$, that is,

$$M_h(\bar{Z}_2) = \frac{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-2}{2}\right)\Gamma\left(\frac{N-n}{2}+h\right)\Gamma\left(\frac{N-n+1}{2}+h\right)}{\Gamma\left(\frac{N-n}{2}\right)\Gamma\left(\frac{N-n+1}{2}\right)\Gamma\left(\frac{N-1}{2}+h\right)\Gamma\left(\frac{N-2}{2}+h\right)},$$

and hence, from (5), we find the distribution of $\bar{Z}_2$ to be

$$f_2(\bar{Z}_2) = \frac{\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-2}{2}\right)\bar{Z}_2^{\frac{N-n-2}{2}}(1-\bar{Z}_2)^{n-3}}{\Gamma\left(\frac{N-n+1}{2}\right)\Gamma\left(\frac{N-n}{2}\right)\Gamma(n-2)} F\left[\frac{n-3}{2}, \frac{n-2}{2}, n-2, 1-\bar{Z}_2\right].$$

It can be shown that $\bar{Z}$ is the λ-criterion appropriate to testing the hypothesis that the n variables in the population fall into k groups, in each of which the variates may be inter-correlated, but such that no variate in one group is correlated with any variate in another.

The practical application of the criteria developed in this paper must be left for further discussion.

MISCELLANEA.

On a Method of proceeding from partial Cell Frequencies to Ordinates and to total Cell Frequencies in the case of a bivariate Frequency Surface.

BY JACQUES CHAPELIN, D.Sc.

IT may happen that to define a bivariate population, the whole population in every cell is not observed, but only the population in a partial cell.

For instance, in order to get a rough evaluation of the amount of wood in a forest, foresters used to divide the area of the forest into rectangular cells, and in each of the rectangles to measure only the volume of the trees falling inside a partial domain, sometimes a strip parallel to one of the sides of the rectangle, coaxial with it and with breadth one-tenth or one-twentieth of the other side. Calling ϕ the (measured) volume corresponding to such a strip or partial cell, it is required to find a plausible value for the volume f inside the rectangle or total cell from which, by addition, a plausible value for the volume of the forest may be deduced. An easy solution would be to use a simple rule of proportionality: the volume of wood in a total cell would be taken equal to the volume in the corresponding partial cell multiplied by the ratio of the area of a total cell to a partial cell. This supposes that the z-ordinate corresponding to the ideal density surface is satisfactorily represented by the z-ordinate of a hyperbolic paraboloid. A more refined solution would be to use Pearson's interpolation surface of the fourth order*. The aim of this paper is to obtain the necessary formulae: in the general case, the value of f for a total rectangle R is a linear and homogeneous function of the values of the nine ϕ's corresponding to this rectangle R and to the eight rectangles adjoining R, and this linear form is defined by the first line of Table III.

To take another example, let us suppose we have a population of N living animals, classified into classes, according to a character X, to which corresponds a first variate x. We wish to study the lethal dose y of a drug, according to the character X. For reasons of economy, we do not want, either to spend too much of this drug or to kill the whole population. Then, we decide to try the drug on one-fifth of the population. We divide each class $(x-\frac{1}{2},\ x+\frac{1}{2})$ into five equal parts, and we experiment only on the animals belonging to the middle class. Thus, we are led to partial cells with breadth $\frac{1}{5}$ and height 1, giving a two-dimensional set of ϕ's from which we shall have to deduce the f's, in order to be able to build up the usual correlation Table. To that effect, we could also use the formulae defined by the Table III of this paper.

We shall suppose that the basic net is a system of squares with unit sides and that the system of partial cells consists of rectangles concentric and coaxial with the basic square cells. We shall use the ordinary Pearsonian interpolation formula corresponding to the mid-panel central difference formula up to and including second order differences. The sides of any of the rectangular cells will be α and β, and we shall call ϕ_{00}, ϕ_{01}, ... the nine observed frequencies in nine partial rectangular cells, according to the usual Pearsonian scheme. If $\alpha=\beta=1$, the quantities ϕ_{tt} reduce to the usual frequencies f_{tt}. The problem is to find the nine total frequencies f_{tt} when the nine partial frequencies ϕ_{tt} are known.

* Cf. *Biometrika*, Vol. XVII. p. 312, or *Tables for Statisticians and Biometricians*, Part II, p. xiii.

We have immediately the following integrals:

$$\int_{-\frac{1}{2}\alpha}^{+\frac{1}{2}\alpha}(1-x^2)\,dx=\alpha-\frac{\alpha^3}{12},\quad \int_{-\frac{1}{2}\alpha}^{+\frac{1}{2}\alpha}x\,(1-x)\,dx=-\frac{\alpha^3}{12},$$

$$\int_{1-\frac{1}{2}\alpha}^{1+\frac{1}{2}\alpha}x\,(1-x)\,dx=-\frac{\alpha^3}{12},\quad \int_{1-\frac{1}{2}\alpha}^{1+\frac{1}{2}\alpha}(1-x^2)\,dx=-\frac{\alpha^3}{12},\quad \int_{1-\frac{1}{2}\alpha}^{1+\frac{1}{2}\alpha}x\,(1+x)\,dx=2\alpha+\frac{\alpha^3}{12},$$

from which we deduce at once Table I leading to the expressions of the ϕ's as functions of the ordinates z. From this table, we can write the equations

$$\frac{576}{\alpha\beta}\phi_{00}=4\,(12-\alpha^2)\,(12-\beta^2)\,z_{00}+\alpha^2\beta^2 z_{11}+\ldots,\ \ldots.$$

The resolution of this system of nine linear equations leads to Table II, from which we can write the equations

$$576\alpha\beta z_{00}=4\,(12+\alpha^2)\,(12+\beta^2)\,\phi_{00}+\alpha^2\beta^2\phi_{11}+\ldots,\ \ldots.$$

At last, eliminating the ordinates z_{ik} between these nine equations and Pearson's formulae (*Biometrika*, Vol. XVII. p. 312, or *Tables for Statisticians and Biometricians*, Part II, p. xiv, formulae (α) to (ι)), we obtain the formulae defined by Table III. They would read

$$f_{00}=4\,(11+\alpha^2)\,(11+\beta^2)\,\phi_{00}+(1-\alpha^2)\,(1-\beta^2)\,\phi_{11}+\ldots,\ \ldots.$$

It should be noticed that we obtain Pearson's formulae (*loc. cit.*) by supposing that, in Table I, $\alpha=\beta=1$, or that, in Table III, α and β tend to zero, and that $\dfrac{\phi_{ik}}{\alpha\beta}$ tends to z_{ik}. Similarly, Table II should lead to the other set of Pearson's formulae (*Biometrika*, Vol. XVII. p. 313, or *Tables for Statisticians and Biometricians*, Part II, p. xiv, formulae (α') to (ι')), by putting $\alpha=\beta=1$. As this is not so, these formulae should be replaced by the following*:

$$\begin{aligned}
576z_{00}&=676f_{00}+f_{11}+f_{1-1}+f_{-11}+f_{-1-1}-26\,(f_{01}+f_{0-1}+f_{10}+f_{-10}),\\
576z_{11}&=4f_{00}+529f_{11}+f_{-1-1}-23\,(f_{1-1}+f_{-11})+46\,(f_{01}+f_{10})-2\,(f_{0-1}+f_{-10}),\\
576z_{1-1}&=4f_{00}+529f_{1-1}+f_{-11}-23\,(f_{11}+f_{-1-1})+46\,(f_{0-1}+f_{10})-2\,(f_{01}+f_{-10}),\\
576z_{-11}&=4f_{00}+529f_{-11}+f_{1-1}-23\,(f_{11}+f_{-1-1})+46\,(f_{01}+f_{-10})-2\,(f_{0-1}+f_{10}),\\
576z_{-1-1}&=4f_{00}+529f_{-1-1}+f_{11}-23\,(f_{1-1}+f_{-11})+46\,(f_{0-1}+f_{-10})-2\,(f_{01}+f_{10}),\\
576z_{01}&=52f_{00}+f_{-1-1}+f_{1-1}-23\,(f_{11}+f_{-11})+598f_{01}-26f_{0-1}-2\,(f_{10}+f_{-10}),\\
576z_{0-1}&=52f_{00}+f_{11}+f_{-11}-23\,(f_{1-1}+f_{-1-1})+598f_{0-1}-26f_{01}-2\,(f_{10}+f_{-10}),\\
576z_{10}&=52f_{00}+f_{-11}+f_{-1-1}-23\,(f_{11}+f_{1-1})+598f_{10}-26f_{-10}-2\,(f_{01}+f_{0-1}),\\
576z_{-10}&=52f_{00}+f_{11}+f_{1-1}-23\,(f_{-11}+f_{-1-1})+598f_{-10}-26f_{10}-2\,(f_{01}+f_{0-1}).
\end{aligned}$$

Lastly, if the basic cells are rectangles with sides h, k, and if the partial cells are rectangles with sides h', k', concentric and coaxial to the rectangles of the basic cells, the right-hand sides of the equations deduced from Table I should be multiplied by hk, and α and β defined by the relations $h'=\alpha h$, $k'=\beta k$, the right-hand sides of the equations deduced from Table II should be divided by hk, and the right-hand sides of the equations deduced from Table III should remain unchanged.

* [I am extremely obliged to Dr Chapelin for his correction of my formulae. K. P.]

TABLE I.

$576\alpha\beta$	z_{00}	z_{11}	z_{1-1}	z_{-11}	z_{-1-1}	z_{01}	z_{0-1}	z_{10}	z_{-10}
ϕ_{00}	$4(12-\alpha^2)(12-\beta^2)$	$\alpha^2\beta^2$	$\alpha^2\beta^2$	$\alpha^2\beta^2$	$\alpha^2\beta^2$	$2(12-\alpha^2)\beta^2$	$2(12-\alpha^2)\beta^2$	$2\alpha^2(12-\beta^2)$	$2\alpha^2(12-\beta^2)$
ϕ_{11}	$4\alpha^2\beta^2$	$(24+\alpha^2)(24+\beta^2)$	$(24+\alpha^2)\beta^2$	$\alpha^2(24+\beta^2)$	$\alpha^2\beta^2$	$-2\alpha^2(24+\beta^2)$	$-2\alpha^2\beta^2$	$-2(24+\alpha^2)\beta^2$	$-2\alpha^2\beta^2$
ϕ_{1-1}	$4\alpha^2\beta^2$	$(24+\alpha^2)\beta^2$	$(24+\alpha^2)(24+\beta^2)$	$\alpha^2\beta^2$	$\alpha^2(24+\beta^2)$	$-2\alpha^2\beta^2$	$-2\alpha^2(24+\beta^2)$	$-2(24+\alpha^2)\beta^2$	$-2\alpha^2\beta^2$
ϕ_{-11}	$4\alpha^2\beta^2$	$\alpha^2(24+\beta^2)$	$\alpha^2\beta^2$	$(24+\alpha^2)(24+\beta^2)$	$(24+\alpha^2)\beta^2$	$-2\alpha^2(24+\beta^2)$	$-2\alpha^2\beta^2$	$-2\alpha^2\beta^2$	$-2(24+\alpha^2)\beta^2$
ϕ_{-1-1}	$4\alpha^2\beta^2$	$\alpha^2\beta^2$	$\alpha^2(24+\beta^2)$	$(24+\alpha^2)\beta^2$	$(24+\alpha^2)(24+\beta^2)$	$-2\alpha^2\beta^2$	$-2\alpha^2(24+\beta^2)$	$-2\alpha^2\beta^2$	$-2(24+\alpha^2)\beta^2$
ϕ_{01}	$-4(12-\alpha^2)\beta^2$	$\alpha^2(24+\beta^2)$	$\alpha^2\beta^2$	$\alpha^2(24+\beta^2)$	$\alpha^2\beta^2$	$2(12-\alpha^2)(24+\beta^2)$	$2(12-\alpha^2)\beta^2$	$-2\alpha^2\beta^2$	$-2\alpha^2\beta^2$
ϕ_{0-1}	$-4(12-\alpha^2)\beta^2$	$\alpha^2\beta^2$	$\alpha^2(24+\beta^2)$	$\alpha^2\beta^2$	$\alpha^2(24+\beta^2)$	$2(12-\alpha^2)\beta^2$	$2(12-\alpha^2)(24+\beta^2)$	$-2\alpha^2\beta^2$	$-2\alpha^2\beta^2$
ϕ_{10}	$-4\alpha^2(12-\beta^2)$	$(24+\alpha^2)\beta^2$	$(24+\alpha^2)\beta^2$	$\alpha^2\beta^2$	$\alpha^2\beta^2$	$-2\alpha^2\beta^2$	$-2\alpha^2\beta^2$	$2(24+\alpha^2)(12-\beta^2)$	$2\alpha^2(12-\beta^2)$
ϕ_{-10}	$-4\alpha^2(12-\beta^2)$	$\alpha^2\beta^2$	$\alpha^2\beta^2$	$(24+\alpha^2)\beta^2$	$(24+\alpha^2)\beta^2$	$-2\alpha^2\beta^2$	$-2\alpha^2\beta^2$	$2\alpha^2(12-\beta^2)$	$2(24+\alpha^2)(12-\beta^2)$

TABLE II.

$576\alpha\beta$	ϕ_{00}	ϕ_{11}	ϕ_{1-1}	ϕ_{-11}	ϕ_{-1-1}	ϕ_{01}	ϕ_{0-1}	ϕ_{10}	ϕ_{-10}
z_{00}	$4(12+\alpha^2)(12+\beta^2)$	$\alpha^2\beta^2$	$\alpha^2\beta^2$	$\alpha^2\beta^2$	$\alpha^2\beta^2$	$-2(12+\alpha^2)\beta^2$	$-2(12+\alpha^2)\beta^2$	$-2\alpha^2(12+\beta^2)$	$-2\alpha^2(12+\beta^2)$
z_{11}	$4\alpha^2\beta^2$	$(24-\alpha^2)(24-\beta^2)$	$-(24-\alpha^2)\beta^2$	$-\alpha^2(24-\beta^2)$	$\alpha^2\beta^2$	$2\alpha^2(24-\beta^2)$	$-2\alpha^2\beta^2$	$2(24-\alpha^2)\beta^2$	$-2\alpha^2\beta^2$
z_{1-1}	$4\alpha^2\beta^2$	$-(24-\alpha^2)\beta^2$	$(24-\alpha^2)(24-\beta^2)$	$\alpha^2\beta^2$	$-\alpha^2(24-\beta^2)$	$-2\alpha^2\beta^2$	$2\alpha^2(24-\beta^2)$	$2(24-\alpha^2)\beta^2$	$-2\alpha^2\beta^2$
z_{-11}	$4\alpha^2\beta^2$	$-\alpha^2(24-\beta^2)$	$\alpha^2\beta^2$	$(24-\alpha^2)(24-\beta^2)$	$-(24-\alpha^2)\beta^2$	$2\alpha^2(24-\beta^2)$	$-2\alpha^2\beta^2$	$-2\alpha^2\beta^2$	$2(24-\alpha^2)\beta^2$
z_{-1-1}	$4\alpha^2\beta^2$	$\alpha^2\beta^2$	$-\alpha^2(24-\beta^2)$	$-(24-\alpha^2)\beta^2$	$(24-\alpha^2)(24-\beta^2)$	$-2\alpha^2\beta^2$	$2\alpha^2(24-\beta^2)$	$-2\alpha^2\beta^2$	$2(24-\alpha^2)\beta^2$
z_{01}	$4(12+\alpha^2)\beta^2$	$-\alpha^2(24-\beta^2)$	$\alpha^2\beta^2$	$-\alpha^2(24-\beta^2)$	$\alpha^2\beta^2$	$2(12+\alpha^2)(24-\beta^2)$	$-2(12+\alpha^2)\beta^2$	$-2\alpha^2\beta^2$	$-2\alpha^2\beta^2$
z_{0-1}	$4(12+\alpha^2)\beta^2$	$\alpha^2\beta^2$	$-\alpha^2(24-\beta^2)$	$\alpha^2\beta^2$	$-\alpha^2(24-\beta^2)$	$-2(12+\alpha^2)\beta^2$	$2(12+\alpha^2)(24-\beta^2)$	$-2\alpha^2\beta^2$	$-2\alpha^2\beta^2$
z_{10}	$4\alpha^2(12+\beta^2)$	$-(24-\alpha^2)\beta^2$	$-(24-\alpha^2)\beta^2$	$\alpha^2\beta^2$	$\alpha^2\beta^2$	$-2\alpha^2\beta^2$	$-2\alpha^2\beta^2$	$2(24-\alpha^2)(12+\beta^2)$	$-2\alpha^2(12+\beta^2)$
z_{-10}	$4\alpha^2(12+\beta^2)$	$\alpha^2\beta^2$	$\alpha^2\beta^2$	$-(24-\alpha^2)\beta^2$	$-(24-\alpha^2)\beta^2$	$-2\alpha^2\beta^2$	$-2\alpha^2\beta^2$	$-2\alpha^2(12+\beta^2)$	$2(24-\alpha^2)(12+\beta^2)$

TABLE III.

$576\alpha\beta$	ϕ_{00}	ϕ_{11}	ϕ_{1-1}	ϕ_{-11}	ϕ_{-1-1}	ϕ_{01}	ϕ_{0-1}	ϕ_{10}	ϕ_{-10}
f_{00}	$4(11+\alpha^2)(11+\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$2(11+\alpha^2)(1-\beta^2)$	$2(11+\alpha^2)(1-\beta^2)$	$2(1-\alpha^2)(11+\beta^2)$	$2(1-\alpha^2)(11+\beta^2)$
f_{11}	$4(1-\alpha^2)(1-\beta^2)$	$(25-\alpha^2)(25-\beta^2)$	$(25-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(25-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(25-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(25-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$
f_{1-1}	$4(1-\alpha^2)(1-\beta^2)$	$(25-\alpha^2)(1-\beta^2)$	$(25-\alpha^2)(25-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(25-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(25-\beta^2)$	$-2(25-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$
f_{-11}	$4(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(25-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(25-\alpha^2)(25-\beta^2)$	$(25-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(25-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(25-\alpha^2)(1-\beta^2)$
f_{-1-1}	$4(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(25-\beta^2)$	$(25-\alpha^2)(1-\beta^2)$	$(25-\alpha^2)(25-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(25-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(25-\alpha^2)(1-\beta^2)$
f_{01}	$-4(11+\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(25-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(25-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$2(11+\alpha^2)(25-\beta^2)$	$2(11+\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$
f_{0-1}	$-4(11+\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(25-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(25-\beta^2)$	$2(11+\alpha^2)(1-\beta^2)$	$2(11+\alpha^2)(25-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$
f_{10}	$-4(1-\alpha^2)(11+\beta^2)$	$(25-\alpha^2)(1-\beta^2)$	$(25-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$2(25-\alpha^2)(11+\beta^2)$	$-2(1-\alpha^2)(11+\beta^2)$
f_{-10}	$-4(1-\alpha^2)(11+\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(1-\alpha^2)(1-\beta^2)$	$(25-\alpha^2)(1-\beta^2)$	$(25-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(1-\beta^2)$	$-2(1-\alpha^2)(11+\beta^2)$	$2(25-\alpha^2)(11+\beta^2)$

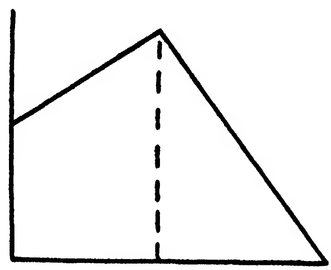
On the Betas of Quadrilateral Distributions.

By OWEN L. DAVIES.

The Pearson Type curves are generated by a differential equation of the form

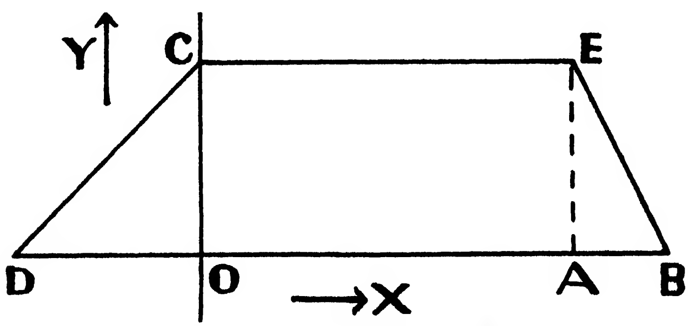
$$\frac{1}{y}\frac{dy}{dx}=\frac{x+a}{b_0+b_1x+b_2x^2} \quad \text{...(1).}$$

These curves cover the whole range of distributions which are likely to be encountered in the field of practical statistics. There are, however, several possible distributions not included in (1) which, although rare in actual experience, have more than mere mathematical interest. Such, for example, is the trapezium or triangle and in fact all curves for which $\frac{dy}{dx}$ is not continuous at all points. The most interesting among these are the trapezia and quadrilaterals with one point of discontinuity for $\frac{dy}{dx}$, namely, figures of the type



Triangular, rectangular and linear distributions are all particular cases of these.

I. *Distributions following a Trapezium Law.*



$OD=d, \quad OA=a, \quad OB=b, \quad OC=c,$

Range $=DB=d+b.$

The equations of the lines DC and EB are respectively

$$y_1=c\left(1+\frac{x}{d}\right),$$

$$y_2=\frac{c}{b-a}(b-x).$$

Let M_s' be the sth moment of the whole figure about the vertical through O, then

$$M_s' = \left\{\int_a^b y_2 x^s dx + c\int_0^a x^s dx\right\} + \int_0^{-d} y_1 x^s dx$$

$$= \frac{c}{(s+1)(s+2)}\left[\left\{\frac{b^{s+2}-a^{s+2}}{b-a}\right\} + (-)^s d^{s+1}\right].$$

Let

$$\left.\begin{aligned} b &= r\cos\theta, \\ a &= r\sin\theta, \qquad \lambda = \cos\theta\sin\theta \\ d &= r\rho, \end{aligned}\right\},$$

then

$$M_s' = \frac{cr^{s+1}}{(s+1)(s+2)}\left[\frac{\cos^{s+2}\theta - \sin^{s+2}\theta}{\cos\theta - \sin\theta} + (-)^s \rho^{s+1}\right],$$

and in particular $\quad M_0' = \frac{cr}{2}[(\cos\theta + \sin\theta) + \rho] = N$ the total frequency.

Now $M_s' = N\mu_s'$, where μ_s' is the sth moment coefficient about O and, therefore, on simplification the first four moment coefficients become

$$\mu_1' = \frac{r}{3}\,\frac{(1+\cos\theta\sin\theta) - \rho^2}{(\cos\theta+\sin\theta)+\rho},$$

$$\mu_2' = \frac{r^2}{6}\,\frac{(\cos\theta+\sin\theta)+\rho^3}{(\cos\theta+\sin\theta)+\rho},$$

$$\mu_3' = \frac{r^3}{10}\,\frac{(1+\cos\theta\sin\theta - \cos^2\theta\sin^2\theta) - \rho^4}{(\cos\theta+\sin\theta)+\rho},$$

$$\mu_4' = \frac{r^4}{15}\,\frac{(\cos\theta+\sin\theta)(1-\cos^2\theta\sin^2\theta)+\rho^5}{(\cos\theta+\sin\theta)+\rho}.$$

Referring these moments to the mean,

$$\mu_2 = \frac{r^2}{18(\rho+\epsilon)^2}[(1+2\lambda-2\lambda^2)+3\rho\epsilon+4\rho^2(1+\lambda)+3\rho^3\epsilon+\rho^4],$$

$$\begin{aligned}\mu_3 = \frac{r^3}{270(\rho+\epsilon)^3}&[(2+6\lambda-3\lambda^2-34\lambda^3)+9\rho\epsilon(1+\lambda-6\lambda^2)+3\rho^2(4-\lambda-29\lambda^2) \\ &-45\rho^3\epsilon\lambda-\rho^4(12+39\lambda)-9\rho^5\epsilon-2\rho^6],\end{aligned}$$

$$\begin{aligned}\mu_4 = \frac{2r^4}{270(\rho+\epsilon)^4}&[(1+\lambda)^2(1+2\lambda-5\lambda^2)+6\rho\epsilon(1+\lambda)(1+\lambda-3\lambda^2) \\ &+\rho^2(17+42\lambda-9\lambda^2-52\lambda^3)+6\rho^3\epsilon(5+6\lambda-5\lambda^2)+3\rho^4(12+24\lambda+\lambda^2) \\ &+6\rho^5\epsilon(5+4\lambda)+\rho^7(17+26\lambda)+6\rho^7\epsilon+\rho^8],\end{aligned}$$

where

$$\lambda = \cos\theta\sin\theta,$$

$$\epsilon = \cos\theta+\sin\theta = (1+2\lambda)^{\frac{1}{2}}.$$

The first two β's, $\qquad \beta_1 = \frac{\mu_3^2}{\mu_2^3}, \quad \beta_2 = \frac{\mu_4}{\mu_2^2},$

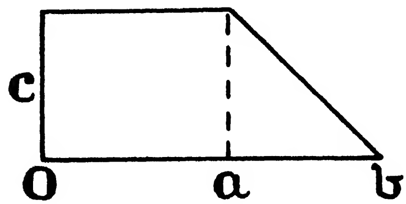
of distributions following a trapezium law will thus depend on two parameters $\lambda = \cos\theta\sin\theta$ and ρ. Now $b \geqslant a$, b and a both positive; consequently, λ is always positive and less than or equal to $\frac{1}{2}$. Moreover, all distributions giving rise to different β's are covered if we take $d \leqslant b$, i.e.

$$O \leqslant \rho \leqslant (\cos\theta - \sin\theta) \leqslant 1.$$

By allowing λ and ρ to vary within the above limits, β_1 and β_2 will be seen to trace out an area of finite extent on the β_1, β_2 plane. The limits to this area may be found by investigating the β_1, β_2 lines of particular subcases.

(i) $d=0$, *i.e.* $\rho=0$.

These distributions will correspond to figures of the following type:



Their moments and β's may be obtained by putting $\rho=0$ in the general relations above. These give

$$\mu_1'=\frac{r}{3}\frac{1+\lambda}{\sqrt{1+2\lambda}},$$

$$\mu_2=\frac{r^2}{18}\frac{1}{(1+2\lambda)}(1+2\lambda-2\lambda^2),$$

$$1 \qquad (2+6\lambda-3\lambda^2-34\lambda^3),$$

$$\mu_4=\frac{2r^4}{270}\frac{1}{(1+2\lambda)^2}(1+\lambda)^2(1+2\lambda-5\lambda^2),$$

whence

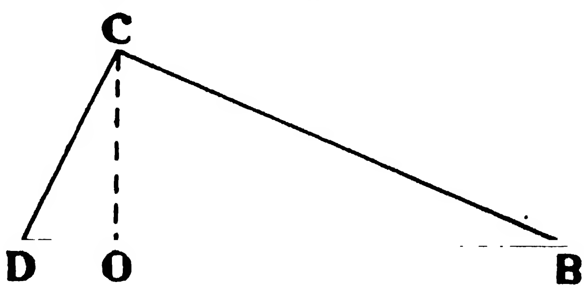
$$\beta_1=\frac{8}{100}\frac{(2+6\lambda-3\lambda^2-34\lambda^3)^2}{(1+2\lambda-2\lambda^2)^3},$$

$$\beta_2=\frac{24}{10}\frac{(1+\lambda)^2(1+2\lambda-5\lambda^2)}{(1+2\lambda-2\lambda^2)^2} \qquad 0\leqslant\lambda\leqslant\tfrac{1}{2}.$$

β_1, β_2 of such distributions will trace out a line of finite extent on the β_1, β_2 plane connecting the line point L to the rectangular point R. These are, in fact, limiting cases corresponding respectively to $\lambda=0$ ($a=0$) and $\lambda=\frac{1}{2}$ ($a=b$) (see Fig. 1).

(ii) $a=0$, *i.e.* $\lambda=0$.

The distributions will now be triangular, corresponding to figures of the type

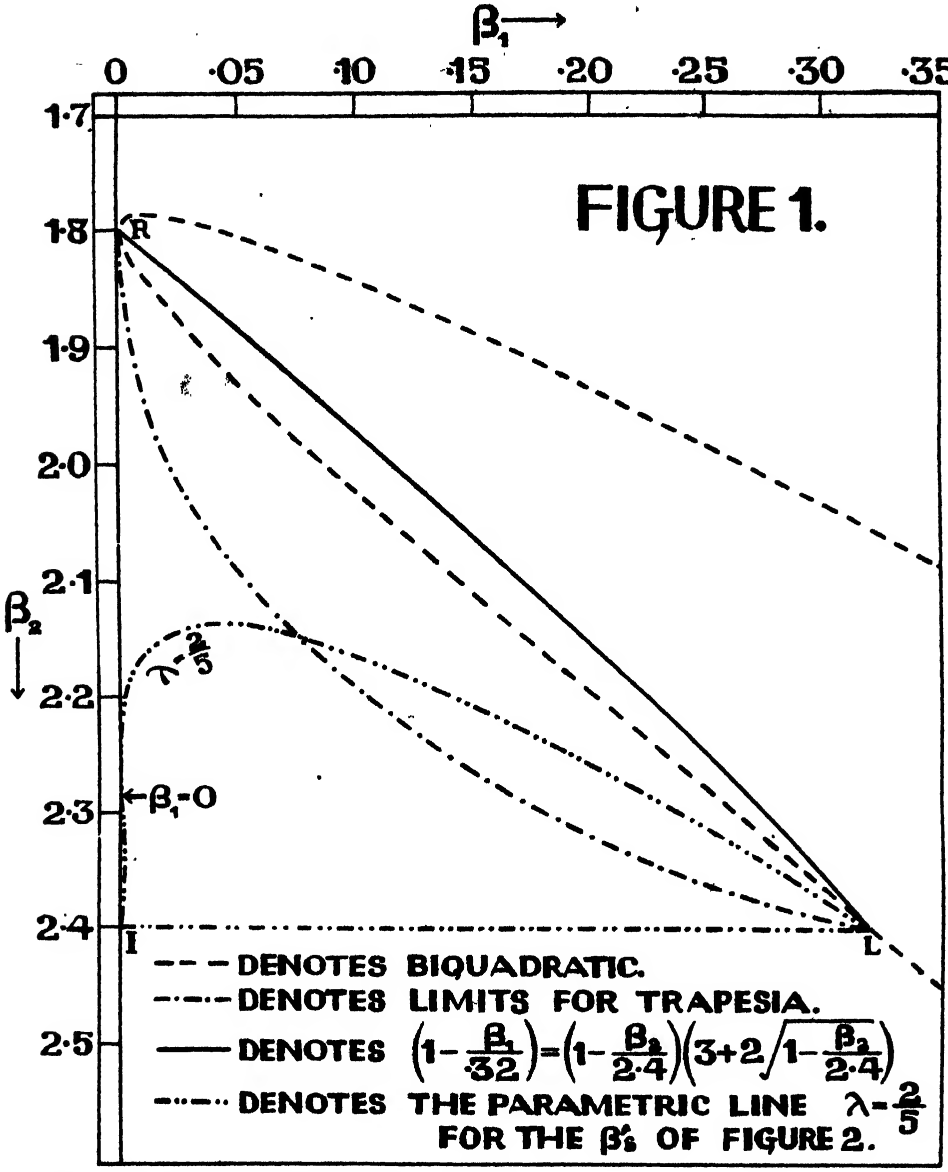


Putting $\lambda=0$ in the general relations, we have

$$\mu_2=\frac{r^2}{18(\rho+1)^2}[1+3\rho+4\rho^2+3\rho^3+\rho^4],$$

$$\mu_3=\frac{r^3}{270(\rho+1)^3}[2+9\rho+12\rho^2-12\rho^4-9\rho^5-2\rho^6],$$

$$\mu_4=\frac{2r^4}{270(\rho+1)^4}[1+6\rho+17\rho^2+30\rho^3+36\rho^4+30\rho^5+17\rho^6+6\rho^7+\rho^8].$$



These may be simplified considerably by writing

$$r\rho = R\cos\phi,$$

$$r = R\sin\phi, \quad \gamma = \cos\phi\sin\phi,$$

whence

$$\mu_2 = \frac{R^2}{18}(1+\gamma),$$

$$\mu_3 = \frac{R^3}{270}(1-2\gamma)^{\frac{1}{2}}(2+5\gamma),$$

$$\mu_4 = \frac{2R^4}{270}(1+\gamma)^2,$$

giving
$$\beta_1=\frac{2}{25}\frac{(1-2\gamma)(2+5\gamma)^2}{(1+\gamma)^3},$$
$$\beta_2=\frac{12}{5} \qquad 0\leqslant\gamma\leqslant\tfrac{1}{2}.$$

β_2 has the same value for all triangular distributions, while β_1 varies between 0 and ·32.

The β_1, β_2 line is, therefore, of finite extent, parallel to the β_1 axis and joins the line point to a point I on the β_2 axis. I corresponds to the symmetrical case, i.e. an isosceles triangle.

(iii) *Symmetrical Case* $(b-a)=d$, *i.e* $\rho=(\cos\theta-\sin\theta)$.

Substituting $\rho=(\cos\theta-\sin\theta)$ in the general relations, we have

$$\mu_2=\frac{r^2}{3(\rho+\epsilon)^2}[(1-\lambda-\lambda^2)+\rho\epsilon(1-\lambda)]$$
$$=\frac{r^2}{24}[(3-4\lambda)+\rho\epsilon],$$
$$\mu_3=0,$$
$$\mu_4=\frac{r^4}{15(\rho+\epsilon)}[(8-16\lambda-18\lambda^2+42\lambda^3-9\lambda^4)+2\rho\epsilon(1-\lambda)(4-4\lambda-5\lambda^2)]$$
$$=\frac{r^4}{480}[(19-44\lambda+18\lambda^2)+(13-20\lambda)\epsilon\rho],$$

whence
$$\beta_1=0,$$
$$\beta_2=\frac{6}{10}\frac{(19-44\lambda+18\lambda^2)+(13-20\lambda)\epsilon\rho}{(5-12\lambda+6\lambda^2)+(3-4\lambda)\epsilon\rho},$$
$$\epsilon\rho=(\cos^2\theta-\sin^2\theta)=(1-4\lambda^2)^{\frac{1}{2}} \qquad 0\leqslant\lambda\leqslant\tfrac{1}{2}.$$

The substitution
$$(1-4\gamma^2)^{\frac{1}{2}}=2\left[\frac{(1-2\lambda)+\epsilon\rho}{(3-4\lambda)+\epsilon\rho}\right]$$
will reduce β_2 to the simple expression
$$\beta_2=\tfrac{12}{5}(1-\gamma^2).$$

When
$$\left.\begin{aligned}\lambda&=0, \text{ then } \gamma=0\\ \lambda&=\tfrac{1}{2}, \text{ then } \gamma=\tfrac{1}{2}\end{aligned}\right\}.$$
and

The limits for γ are, therefore, the same as those for λ, namely $(0, \tfrac{1}{2})$.

β_2 thus varies between 1·8 and 2·4 and the β_1, β_2 line is that part of the β_2 axis lying between the rectangular point R and the isosceles point I.

Differentiating β_1 and β_2 for the general case, we find

$$\left(\frac{\partial\beta_2}{\partial\rho}\right)_{\rho=\lambda=0}=0, \qquad \left(\frac{\partial\beta_2}{\partial\lambda}\right)_{\rho=\lambda=0}=0,$$
$$\left(\frac{\partial\beta_1}{\partial\rho}\right)_{\rho=\lambda=0}=0, \qquad \left(\frac{\partial\beta_1}{\partial\lambda}\right)_{\rho=\lambda=0}=0,$$

and
$$(\beta_2)_{\lambda=\rho=0}=2{\cdot}4, \qquad (\beta_1)_{\lambda=\rho=0}={\cdot}32.$$

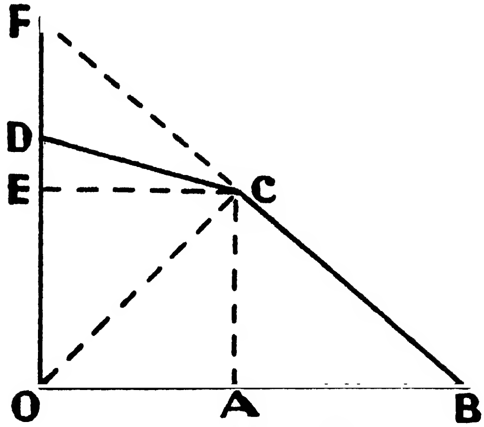
The β_1, β_2 area is, therefore, bounded below by the line $\beta_2=2{\cdot}4$. Moreover, β_1 never exceeds ·32 and β_2 is never less than 1·8. It is fairly evident, then, that the area traced out by β_1 and β_2 of distributions following a trapezium law is bounded by the β_1, β_2 lines of the following three subcases:

(i) triangle,

(ii) symmetrical trapezium,

(iii) the quadrilateral formed by a rectangle and a right triangle (see Fig. 1).

II. *Quadrilaterals with One Point of Discontinuity for dy/dx.*



$OA=a, \quad OB=b, \quad AC=c, \quad ED=d=\rho c.$

The sth moment of the quadrilateral $OBCD$ about O is given by

$$M_s'=\frac{c}{(s+1)(s+2)}\left[\frac{b^{s+2}-a^{s+2}}{b-a}+a^{s+1}\rho\right],$$

and
$$M_0'=\frac{c}{2}\left[(b+a)+a\rho\right].$$

Write
$$b=r\cos\theta,$$
$$a=r\sin\theta, \qquad \lambda=\cos\theta\sin\theta,$$

then the first four moment coefficients about O become

$$\mu_1'=\frac{r}{3}\frac{(1+\lambda)+\rho\sin^2\theta}{(\cos\theta+\sin\theta)+\rho\sin\theta},$$

$$\mu_2'=\frac{r^2}{6}\frac{(\cos\theta+\sin\theta)+\rho\sin^3\theta}{(\cos\theta+\sin\theta)+\rho\sin\theta},$$

$$\mu_3'=\frac{r^3}{10}\frac{(1+\lambda-\lambda^2)+\rho\sin^4\theta}{(\cos\theta+\sin\theta)+\rho\sin\theta},$$

$$\mu_4'=\frac{r^4}{15}\frac{(\cos\theta+\sin\theta)(1-\lambda^2)+\rho\sin^5\theta}{(\cos\theta+\sin\theta)+\rho\sin\theta}.$$

Referring these moments to the centroid vertical,

$$\text{(A)} \quad \mu_2=\frac{r^2}{18(\epsilon+\rho\sin\theta)^2}\left[(1+2\lambda-2\lambda^2)+\rho\{3\lambda(1-\lambda)+\sin^2\theta(2-\lambda)\}+\rho^2\sin^4\theta\right],$$

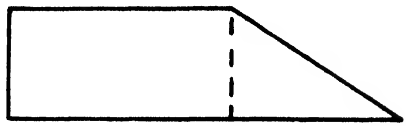
$$\mu_3=\frac{r^3}{270(\epsilon+\rho\sin\theta)^3}\left[(2+6\lambda-3\lambda^2-34\lambda^3)+\rho\{9\lambda(1+3\lambda-7\lambda^2)+\sin^2\theta(6+3\lambda-39\lambda^2)\}\right.$$
$$\left.+3\rho^2\{2\sin^2\theta(1+3\lambda)(1-2\lambda)-\lambda^2(8\lambda-7)\}+2\rho^3\sin^6\theta\right],$$

$$\mu_4=\frac{2r^4}{270(\epsilon+\rho\sin\theta)^4}\left[(1+\lambda)^2(1+2\lambda-5\lambda^2)+\rho\{3\lambda(2+2\lambda-5\lambda^2-5\lambda^3)\right.$$
$$+\sin^2\theta(4+6\lambda-12\lambda^2-5\lambda^3)\}+3\rho^2\{\lambda^2(1-5\lambda^2)+\sin^2\theta(2+4\lambda-7\lambda^2-4\lambda^3)\}$$
$$\left.+\rho^3\{\lambda^2(-4+5\lambda-3\lambda^2)+\sin^2\theta(4+4\lambda-10\lambda^2-7\lambda^3)\}+\rho^4\sin^8\theta\right].$$

Subcases.

(i) $d=0$, *i.e.* $\rho=0$.

The moments and β's reduce for those already found for the figure (p. 500).



(ii) $d=-c$, *i.e.* $\rho=-1$.

This subcase corresponds to triangular distributions, and on substitution the three moments reduce to

$$\mu_2=\frac{r^2}{18}(1-\lambda),$$

$$\mu_3=\frac{r^3}{270}(1+2\lambda)^{\frac{1}{2}}(2-5\lambda),$$

$$\mu_4=\frac{2r^4}{270}(1-\lambda)^2,$$

giving

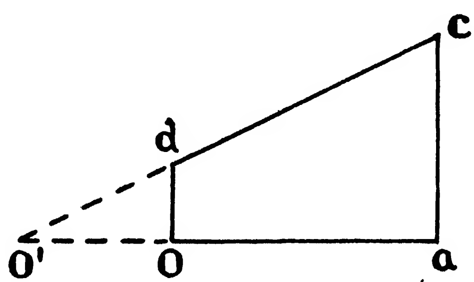
$$\beta_1=\frac{2}{25}\,\frac{(1+2\lambda)(2-5\lambda)^2}{(1-\lambda)^3},$$

$$\beta_2=\frac{12}{5} \qquad 0\leqslant\lambda\leqslant\tfrac{1}{2}.$$

This represents the same line already found for triangular distributions. The λ, however, has a different meaning. The figure is symmetrical when $a=2b$, i.e. $\cos\theta=2\sin\theta$ or $\lambda=\frac{2}{5}$. For this value of λ, $\beta_1=0$ as expected. For $\lambda=0$ or $\frac{1}{2}$ the figure corresponds to a right triangle giving $\beta_2=2{\cdot}4$, $\beta_1={\cdot}32$, the coordinates of the line point.

(iii) $a=b$, *i.e.* $\lambda=\frac{1}{2}$.

The distributions will now be linear and correspond to figures of the following type.



Substituting $\lambda=\frac{1}{2}$ in relations (A), we have

$$\mu_2=\frac{\left(\frac{r}{\sqrt{2}}\right)^2}{18}\,\frac{6+6\rho+\rho^2}{(\rho+2)^2},$$

$$\mu_3=\frac{\left(\frac{r}{\sqrt{2}}\right)^3}{270}\,\frac{\rho(9+9\rho+\rho^2)}{(\rho+2)^3},$$

$$\mu_4=\frac{2\left(\frac{r}{\sqrt{2}}\right)^4}{270}\,\frac{27+54\rho+39\rho^2+12\rho^3+\rho^4}{(\rho+2)^4},$$

giving
$$\beta_1 = \frac{32}{100}\rho^2\frac{(9+9\rho+\rho^2)^2}{(6+6\rho+\rho^2)^3},$$
$$\beta_2 = \frac{12}{5}\frac{27+54\rho+39\rho^2+12\rho^3+\rho^4}{(6+6\rho+\rho^2)^2}.$$

Put
$$(\rho+1)=\frac{\sin\theta}{\cos\theta} \text{ and } \lambda'=\cos\theta\sin\theta,$$

then
$$\beta_1=\frac{32}{100}\frac{(1-2\lambda')(1+7\lambda')^2}{(1+4\lambda')^3},$$
$$\beta_2=\frac{12}{5}\frac{(1+7\lambda')(1+\lambda')}{(1+4\lambda')^2} \qquad 0\leqslant\lambda'\leqslant\tfrac{1}{2}.$$

We may give the following interpretation to λ'. Let $O'O=b$, $O'a=a$ and write
$$b=r\sin\theta,$$
$$a=r\cos\theta, \qquad r^2=a^2+b^2,$$
then λ' will be equal to $\cos\theta\sin\theta$.

The curve* connecting β_1 and β_2 is
$$\left(1-\frac{\beta_1}{\cdot 32}\right)=\left(1-\frac{\beta_2}{2\cdot 4}\right)\left(3+2\sqrt{1-\frac{\beta_2}{2\cdot 4}}\right) \quad\text{......(2).}$$

It passes through the points
$$\beta_1=\cdot 32, \quad \beta_2=2\cdot 4,$$
$$\beta_1=0, \quad \beta_2=1\cdot 8,$$
which are, respectively, the coordinates of the line point L and the rectangular point R.

Clearly, β_2 must be less than or equal to 2·4, and accordingly β_1 must be less than or equal to ·32.

The line (2) may be readily plotted from the parametric equations by allowing λ' to vary between 0 and $\frac{1}{2}$. This line is of finite extent connecting the line point and the rectangular point, and lies entirely within the biquadratic loop which is of fundamental importance in connection with Pearson's Type curves. It is of interest, therefore, to determine how closely a Type I curve
$$y=y_0\left(1+\frac{x}{a_1}\right)^{m_1}\left(1-\frac{x}{a_2}\right)^{m_2} \quad\text{......(3)}$$
will fit a linear distribution.

The best fit is obtained by identifying the first two moments and the range in both cases. For (3) we have†
$$\mu_1'=\frac{m_1'}{m_1'+m_2'},$$
$$\left.\begin{aligned}\sigma^2=\frac{m_1'm_2'}{(m_1'+m_2')^2(m_1'+m_2'+1)} \qquad & m_1'=m_1+1\\ & m_2'=m_2+1\end{aligned}\right\}.$$

Hence
$$\frac{\mu_1'^2}{\sigma^2}=\frac{m_1'}{m_2'}(m_1'+m_2'+1),$$
$$\frac{1}{\mu_1'}=\left(1+\frac{m_2'}{m_1'}\right),$$

and, therefore,
$$m_1'=m_1+1=\frac{(1-\mu_1')\mu_1'^2}{\sigma^2}-\mu_1' \quad\text{......(4),}$$
$$m_2'=m_2+1=\left(\frac{1-\mu_1'}{\mu_1'}\right)(m_1+1) \quad\text{......(5).}$$

* Due to K. Pearson.

† K. Pearson, *Phil. Trans.*, Vol. 186, p. :

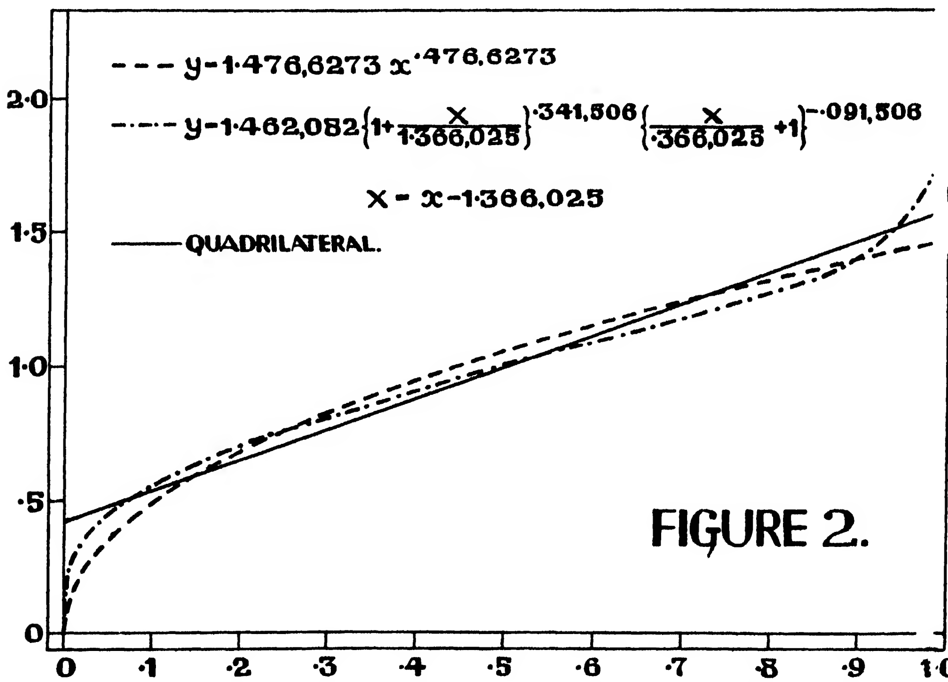
Since $\frac{m_1}{m_2}=\frac{a_1}{a_2}$ and $a_1+a_2=$ total range R, we may readily calculate the constants m_1, m_2, a_1 and a_2. Moreover, the area of the whole curve is equal to the total frequency N. This enables us to find the remaining constant y_0.

Example. $\lambda' = \cdot 25$.

We readily find
$$R = r(\cos\theta - \sin\theta) = r \times \cdot 707{,}1068,$$
$$\frac{\mu_1'}{R} = \cdot 596{,}2251,$$
$$\left(\frac{\sigma}{R}\right)^2 = \cdot 074{,}07407,$$

and the required Type I curve is found to be
$$\left(\frac{y}{N}\right) = 1\cdot 462{,}082\left(1+\frac{x/R}{\cdot 366{,}025}\right)^{-\cdot 091{,}506}\left(1+\frac{x/R}{1\cdot 366{,}025}\right)^{\cdot 341{,}506} \quad\text{.................(6).}$$

How closely this curve fits the linear distribution may be judged from Fig. 2.



Throughout the whole range of variation of λ', the β_1, β_2 line remains fairly close to the lower branch of the biquadratic. We may, therefore, reasonably expect a fairly good fit of the type
$$y = y_0\left(1+\frac{x}{a}\right)^{m_1} \quad \text{(Pearson Type IX curve).}$$
Range $x = -a$ to 0.

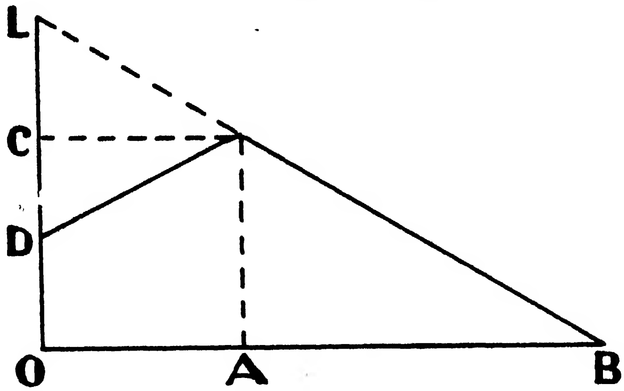
Identifying the range and mean of the two distributions, we have
$$\left(\frac{m_1+1}{m_1+2}\right) = \frac{\mu_1'}{R},$$
from which m_1 may be calculated. y_0 is found by equating the whole area to the total frequency N.

For $\lambda' = \cdot 25$, the curve is found to be

$$\left(\frac{y}{N}\right) = 1{\cdot}476{,}6273\,(1+x/R)^{\cdot 476{,}6273} \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7),$$

which is plotted together with (6) and the line in Fig. 2. It can be seen that (7) is almost as good a fit as the more general curve (6)*.

Let us return now to the general distributions of the type :



The first two β's were found to depend on two parameters λ and ρ. The β's, therefore, trace t an area on the β_1, β_2 plane.

(i) *When the point D varies between O and C*, i.e. $-1 \leqslant \rho \leqslant 0$ and $0 \leqslant \lambda \leqslant \frac{1}{2}$, the β_1, β_2 area is ntical with the one mapped out by the first two β's of trapezia.

(ii) *When D lies between C and L*, i.e. $0 \leqslant \rho \leqslant \dfrac{\sin\theta}{\cos\theta - \sin\theta}$ and $0 \leqslant \lambda \leqslant \frac{1}{2}$, the area traced out by the β's is that part of the plane bounded by the β_1, β_2 lines of

(*a*) linear distributions,

(*b*) distributions represented by quadrilaterals formed by placing together a rectangle and a right triangle.

(iii) *When D lies beyond L*, i.e. $\rho \geqslant \dfrac{\sin\theta}{\cos\theta - \sin\theta} \geqslant 0$, the β_1, β_2 area is bounded by a loop which is the envelope of the lines $\lambda = \text{constant}$ for $\frac{1}{2} \geqslant \lambda \geqslant \frac{3}{8}$, $\rho \geqslant \dfrac{\sin\theta}{\cos\theta - \sin\theta}$.

When $0 \leqslant \lambda \leqslant \frac{3}{8}$, the β_1, β_2 area is bounded by the loop $\lambda = \frac{3}{8}$, $\rho \geqslant 1$, i.e.

$$\beta_1 = \frac{8}{100}\,\frac{(218+510\rho+294\rho^2+2\rho^3)^2}{(37+26\rho+\rho^2)^3},$$

$$\beta_2 = \frac{12}{5}\,\frac{1225+1780\rho+894\rho^2+196\rho^3+\rho^4}{(37+26\rho+\rho^2)^2}.$$

III. *Convex Curves.*

The first two β's of the above quadrilateral distributions occupy a relatively small portion of the β_1, β_2 plane and those which are convex at all points fall within the area bounded by the lines

(i) $\beta_1 = 0$.

(ii) $\beta_2 = 2{\cdot}4$.

(iii) $\left(1 - \dfrac{\beta_1}{\cdot 32}\right) = \left(1 - \dfrac{\beta_2}{2{\cdot}4}\right)\left(3 + 2\sqrt{1 - \dfrac{\beta_2}{2{\cdot}4}}\right)$.

Denote this region by *C*.

* Fig. 2 provides an interesting example of the extent to which equality in the first four moments leads to correspondence in form.

Convex curves are necessarily of limited range and, therefore, of Pearson's curves, those which are convex at all points must have the form

$$\text{(iv)}\quad y = y_0 x^s (1-x)^t.$$

Differentiating twice

$$\frac{d^2y}{dx^2} = y_0 x^{s-2}(1-x)^{t-2}\left[s(s-1) - 2x^s(s+t-1) + x^2(s+t)(s+t-1)\right].$$

This is negative throughout the whole range provided $s+t \leqslant 1$, neither s nor t being negative. This is, therefore, the condition for convexity.

When s or t is zero, the β's of (iv) lie on the lower branch of the biquadratic, and when $s=t$ the curve is symmetrical, giving $\beta_1 = 0$. Furthermore, when $s+t=1$, we have

$$\beta_1 = \frac{16}{25}\left(\frac{9}{\epsilon} - 4\right),$$

$$\beta_2 = \frac{6}{5}\left(\frac{6}{\epsilon} - 1\right) \qquad \epsilon = (2-t)(1+t).$$

Eliminating $\frac{1}{\epsilon}$, we have

$$\text{(v)}\quad 4\beta_2 - 5\beta_1 - 8 = 0,$$

which is a straight line passing through the points $\left.\begin{matrix}\beta_1 = 0\\ \beta_2 = 2\end{matrix}\right\}$ and L.

It is clear then that the region of convexity for the Pearson curves is bounded by

(a) the lower branch of the biquadratic,

(b) the β_2 axis,

(c) $4\beta_2 - 5\beta_1 - 8 = 0$.

This region lies entirely within C.

The normal curve $y = y_0 e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ is convex between its points of inflexion. Let M_s denote the sth moment about the vertical through the origin and $\pm a\sigma$ the points of inflexion, then

$$M_{2s} = 2y_0 \int_0^{a\sigma} e^{-\left(\frac{x}{\sigma}\right)^2} x^{2s}\,dx = y_0 \sigma^{2s+1} \Gamma_{a^2}(s+\tfrac{1}{2}).$$

Hence
$$\frac{\mu_{2s}}{\sigma^{2s}} = \frac{\Gamma_{a^2}(s+\frac{1}{2})}{\Gamma_{a^2}(\frac{1}{2})}.$$

The points of inflexion are given by $\frac{d^2y}{dx^2} = 0$, i.e. $a = \frac{1}{\sqrt{2}}$. Hence

$$\frac{\mu_2}{\sigma^2} = \frac{\Gamma_{\frac{1}{2}}(1{\cdot}5)}{\Gamma_{\frac{1}{2}}({\cdot}5)}; \quad \frac{\mu_4}{\sigma^4} = \frac{\Gamma_{\frac{1}{2}}(2{\cdot}5)}{\Gamma_{\frac{1}{2}}({\cdot}5)},$$

which give
$$\beta_2 = 1{\cdot}941, \quad \beta_1 = 0.$$

For one-half of the curve
$$[illegible]\ \frac{\Gamma_{\frac{1}{2}}\left(\frac{s+1}{[illegible]}\right)}{[illegible]}$$

from which we deduce
$$\left.\begin{matrix}\beta_2 = 1{\cdot}872\\ \beta_1 = \ {\cdot}0273\end{matrix}\right\}.$$

Both these points lie within C.

For the quadrant of the ellipse

$$y=y_0(1-x^2)^{\frac{1}{2}} \qquad (0 \leqslant x \leqslant 1),$$

$$M_s' \text{ (about the origin)} = y_0 \int_0^1 (1-x^2)^{\frac{1}{2}} x^s\,dx$$

$$=\frac{y_0}{2}\frac{\Gamma\left(\frac{3}{2}\right)\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{s+4}{2}\right)}.$$

Hence

$$\mu_s'=\frac{1}{\Gamma(\frac{1}{2})}\frac{\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{s+4}{2}\right)},$$

from which we deduce

$$\left.\begin{aligned}\beta_1&=1\cdot9772\\ \beta_2&=\cdot0554\end{aligned}\right\}.$$

The trigonometrical curve $\qquad y=y_0\cos x \quad \left(-\frac{\pi}{2}\leqslant x\leqslant\frac{\pi}{2}\right)$

is convex. If M_s' denotes the sth moment, about the origin, of one-half of the curve,

$$M_s'=y_0\int_0^{\frac{\pi}{2}}\cos x\, x^s\,dx=M_0'\mu_s',$$

$$M_0'=y_0.$$

After reduction we find

$$\mu_4'=\left(\frac{\pi}{2}\right)^4-12\left(\frac{\pi}{2}\right)^2+24,$$

$$\mu_3'=\left(\frac{\pi}{2}\right)^3-6\left(\frac{\pi}{2}\right)+6,$$

$$\mu_2'=\left(\frac{\pi}{2}\right)^2-2,$$

$$\mu_1'=\left(\frac{\pi}{2}-1\right),$$

whence

$$\beta_2=2\cdot2317,$$
$$\beta_1=\cdot1797.$$

For the complete symmetrical curve we find

$$\beta_2=2\cdot1938,$$
$$\beta_1=0.$$

Both these points fall in the region C.

Finally, the curve $\qquad y=y_0(1-e^{-x}) \qquad (x\geqslant 0)$

is convex at all points. If we take the range (0, 1), we have

$$M_s' \text{ (about the origin)} = y_0\int_0^1 (1-e^{-x})\,x^s\,dx$$

$$=y_0\left\{\left(\frac{1}{s+1}\right)+e^{-1}\right\}-sM'_{s-1}.$$

Evaluating the integrals and referring the moments to the centroid vertical, we have

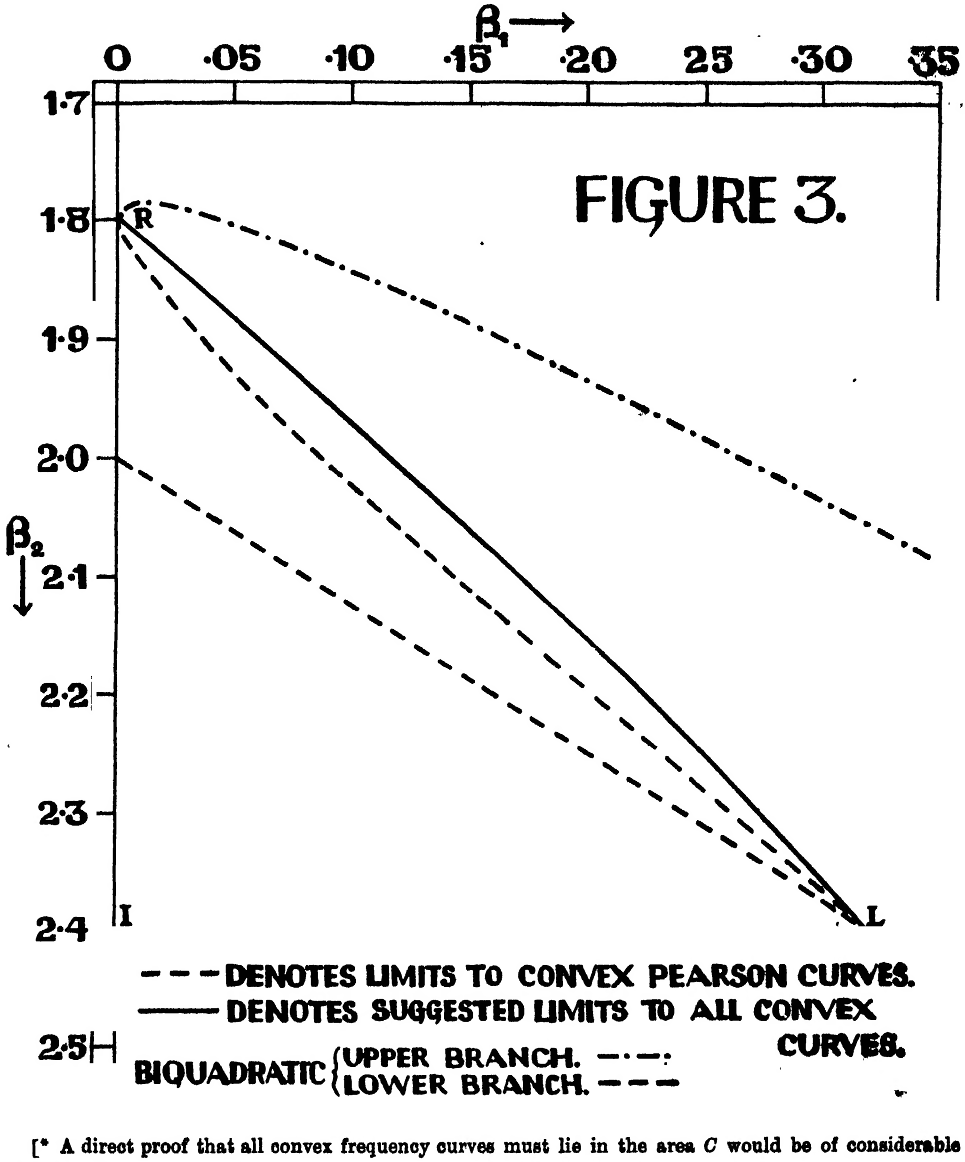
$$\mu_2=\cdot058{,}830,$$
$$\mu_3=-\cdot006{,}427,$$
$$\mu_4=\cdot007{,}716,$$

giving

$$\beta_2=2\cdot2295,$$
$$\beta_1=0\cdot2029,$$

which lies within C.

Consequently, for all convex curves considered above, the first two β's fall in the region C. I cannot conceive of a convex curve which does not give this result, and it seems quite probable that the β's of all convex curves fall in the region C*.



[* A direct proof that all convex frequency curves must lie in the area C would be of considerable interest. Ed.]

CAMBRIDGE: PRINTED BY WALTER LEWIS, M.A., AT THE UNIVERSITY PRESS